



# **ADMI 2024**

Code-a-thon



**TACC**  
TEXAS ADVANCED COMPUTING CENTER



**SGX3**  
Extend. Expand. Exemplify.



# THE PLAN

- Introductions
- What are we doing here?
- What is Software Engineering
- Software Engineering vs Science Gateways vs High Performance Computing (HPC)
- Technology we're going to be using
- Welcome to EUREKA!

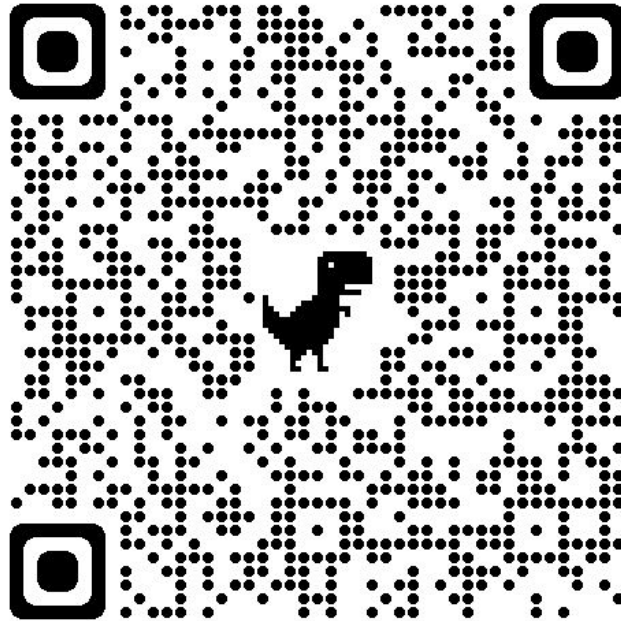






# BUT FIRST!

It's time to Jam.





# THE PLAN

- Introductions
- What are we doing here?
- What is Software Engineering
- Software Engineering vs Science Gateways vs High Performance Computing (HPC)
- Technology we're going to be using
- Welcome to EUREKA!





# WHAT ARE WE DOING HERE?

## Hackathons: A Brief Overview

Hackathons are **intensive, time-bound** events where teams of participants **come together to collaboratively work on solving real-world problems** or **creating innovative software projects.**





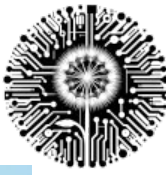


# THE CODE-A-THON!

**Participants** are going to **engage in coding challenges or competitions**, where each **challenge builds on the previous challenge**. These challenges are algorithmic or data structure-related and **each challenge combines together to become a major project**.





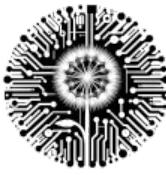


# SOFTWARE ENGINEERING?

- # ● What is Software Engineering
- Requirements Gathering
  - **Software Architecture**
  - **Coding and Programming**
  - Software Testing and Debugging
  - Software Maintenance
  - Software Project Management
  - Software Quality
  - Software Metrics
  - **Software Development Models & Architecture**







# SOFTWARE ENGINEERING vs SCIENCE GATEWAYS vs HIGH PERFORMANCE COMPUTING

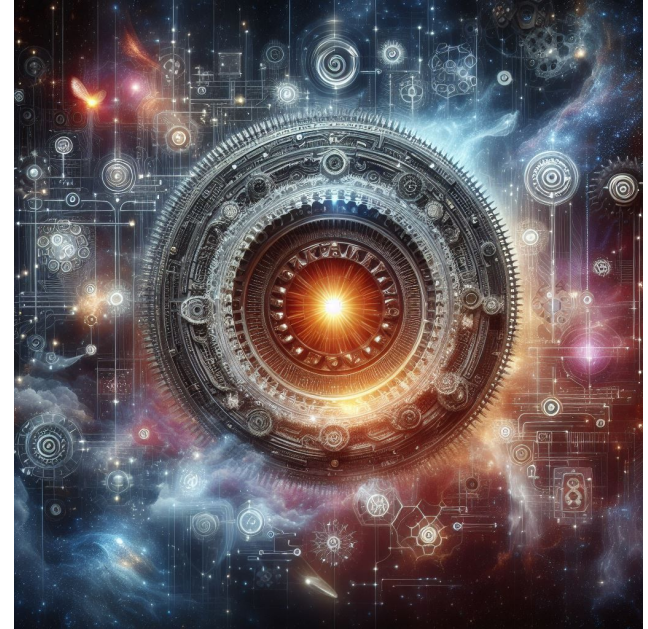




# SOFTWARE ENGINEERING AND SCIENCE GATEWAYS

- What is a Science Gateway?

Science gateways are **user-friendly interfaces** that allow **researchers and educators** to **access advanced resources, tools, applications, and data collections** specific to a science or engineering





# SCIENCE GATEWAYS AND HIGH PERFORMANCE COMPUTING

- So what does this have to do with High Performance Computing?

Science Gateways are connected to High Performance Computing (HPC), they provide a **user-friendly interface to HPC resources**





# "COMPUTING FOR THE ENDLESS FRONTIER"

Dan Stanzione

Executive Director, Texas Advanced Computing Center

Associate Vice President for Research, The University of Texas at Austin



# RAPID GROWTH FROM THEN TO NOW...

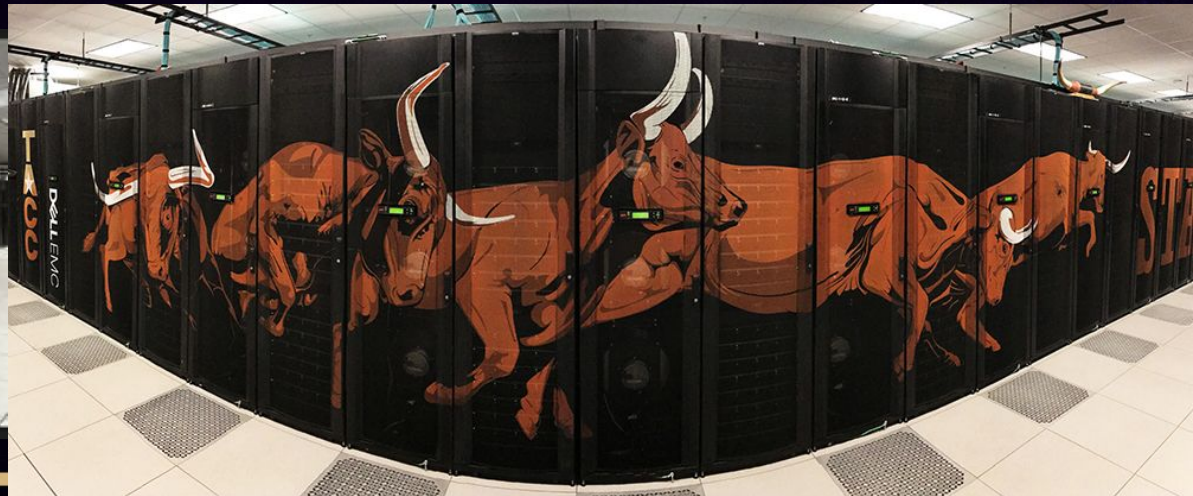
- ▶ 2003 – First Terascale Linux cluster for open science (#26)
- ▶ 2004 – NSF funding to join the Teragrid
- ▶ 2006 – UT System Partnership to provide Lonestar-3 (#12)
- ▶ **2007 - \$59M NSF award – largest in UT history – to deploy Ranger, the world's largest open system (#4)**
- ▶ 2008 – funding for new Vis software and launch of revamped visualization lab.
- ▶ 2009 - \$50M iPlant Collaborative award (largest NSF bioinformatics award) moves a major component to TACC, life sciences group launched.
  - ▶ In 2009, we reached, 65 employees.





# NOW, A WORLD LEADER IN CYBERINFRASTRUCTURE

- ▶ 2010, TACC becomes a core partner (1 of 4) in XSEDE, the TeraGrid Replacement
- ▶ 2012, Stampede replaces Ranger with new \$51.5M NSF Award
- ▶ 2013, iPlant is renewed, expanded to \$100M
- ▶ 2015, Wrangler, first data intensive supercomputer is deployed.
- ▶ 2015, Chameleon cloud is launched
- ▶ 2015, DesignSafe, the cyberinfrastructure for natural hazard engineering, is launched.
- ▶ 2016 Stampede-2 awarded the largest academic system in the United States, 2017-2021.  
(56th fastest super computer in the world)



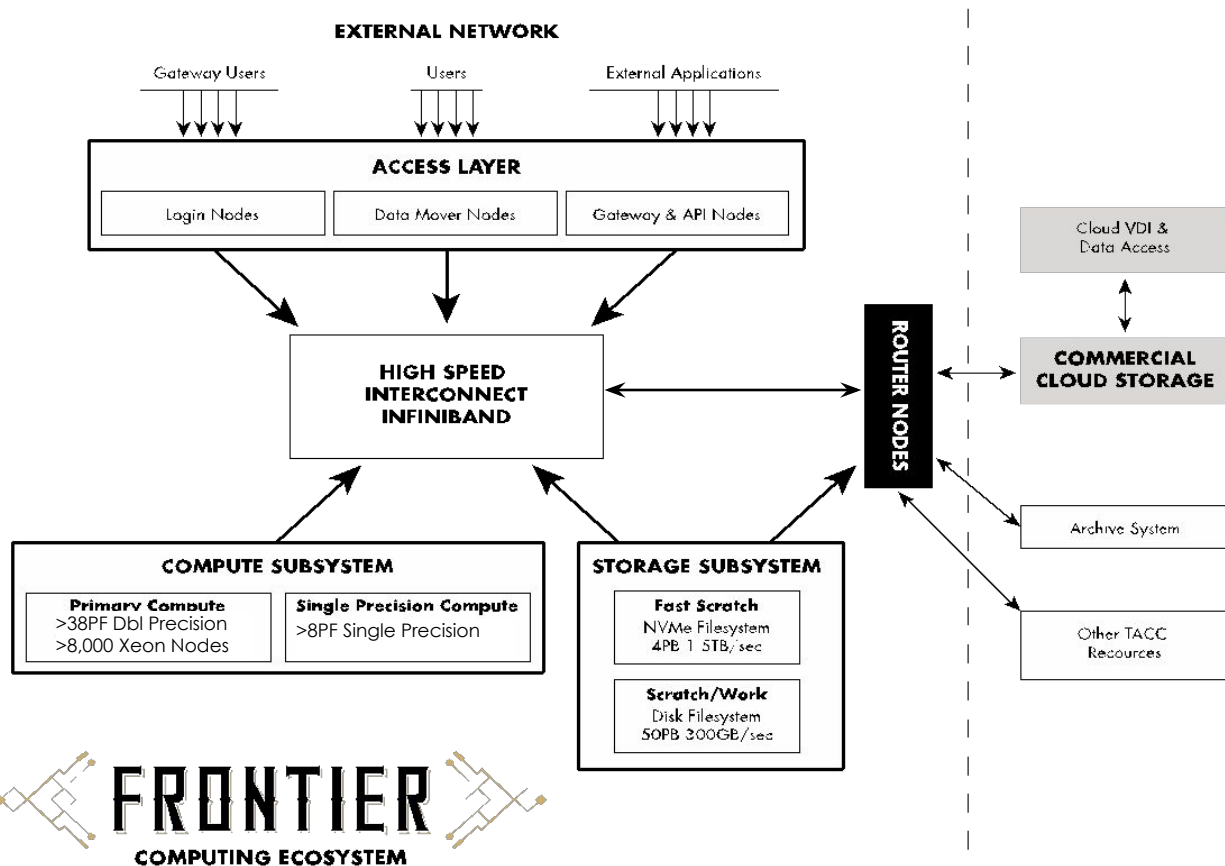


# FRONTERA SYSTEM --- HARDWARE

- ▶ 21st Fastest Supercomputer in the world
  - ▶ #1 for Open Science
- ▶ Primary compute system: DellEMC and Intel
  - ▶ 35-40 PetaFlops Peak Performance (Next Generation Xeon processors)
- ▶ Interconnect: Mellanox HDR and HDR-100 links.
  - ▶ Fat Tree topology, 200Gb/s links between switches.
- ▶ Storage: DataDirect Networks
  - ▶ 50+ PB disk, 3PB of Flash, 1.5TB/sec peak I/O rate.
- ▶ Single Precision Compute Subsystem: Nvidia
- ▶ Front end for data movers, workflow, API



# SYSTEM OVERVIEW









- ▶ **Humphry Davy, Inventor of Electrochemistry, 1812**
- ▶ (Pretty sure he was talking about our machine).

Nothing tends so much to the advancement of knowledge as the application of a new instrument. The native intellectual powers of men in different times are not so much the causes of the different success of their labours, as the peculiar nature of the means and artificial resources in their possession.

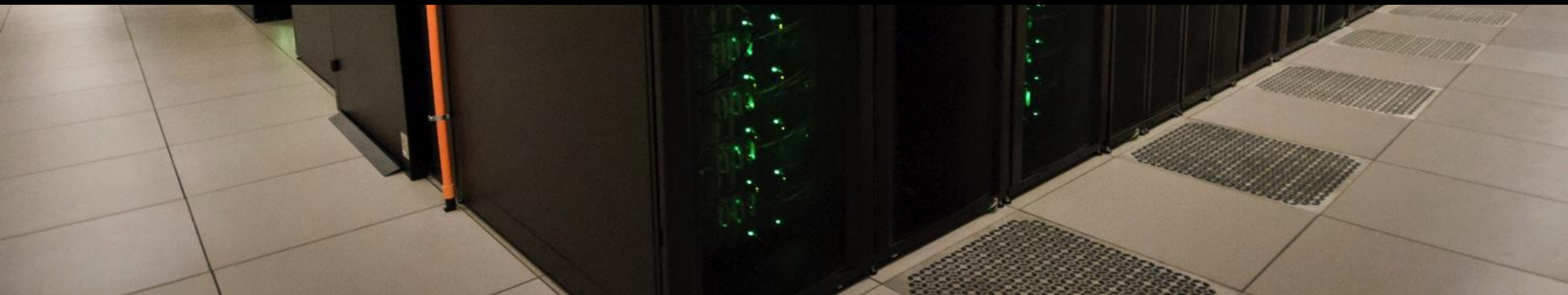
Humphry Davy

PICTUREQUOTES.com

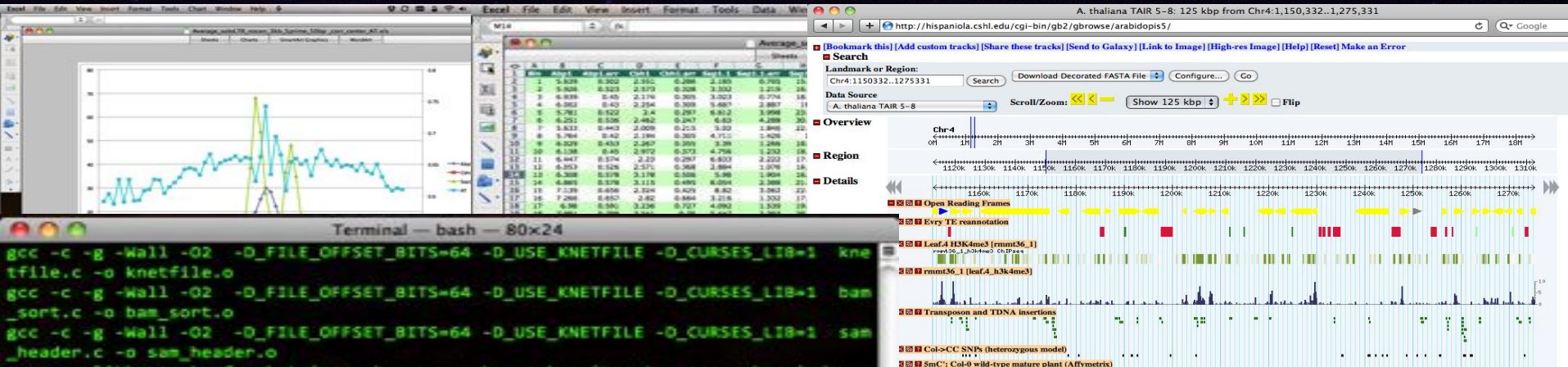
HOW DO WE HELP RESEARCHERS WITH SUCH  
DIVERSE NEEDS AND BACKGROUNDS?



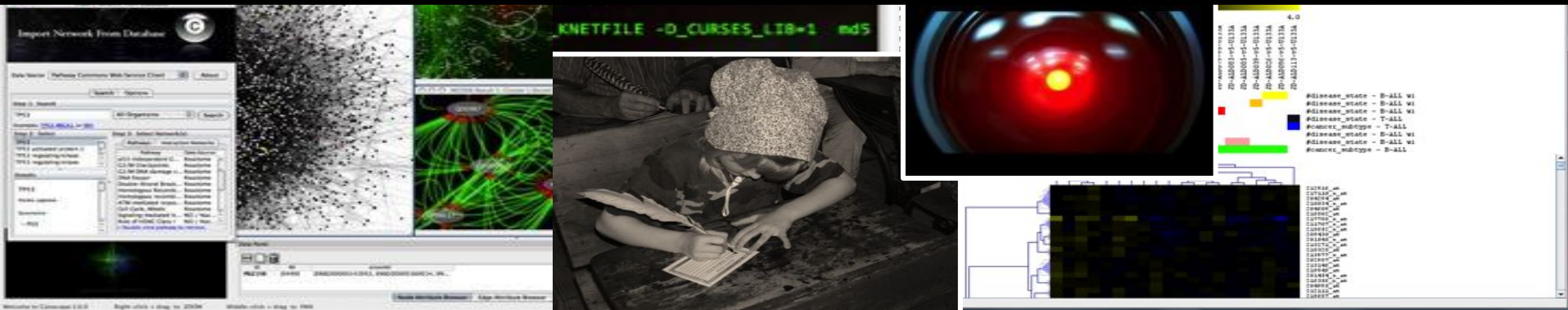
BUILD A MASSIVE STORAGE CLOUD NEXT TO INNOVATIVE, POWERFUL, USABLE  
COMPUTERS AT THE END OF FAST INTERNET PIPES

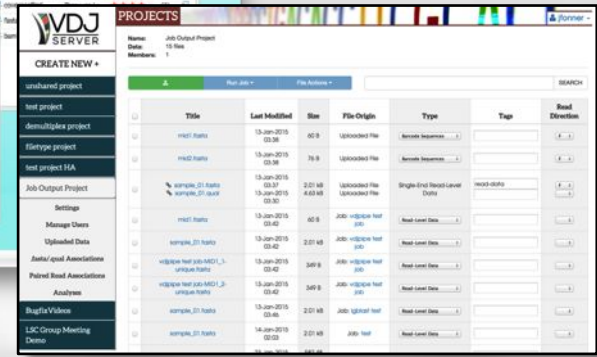
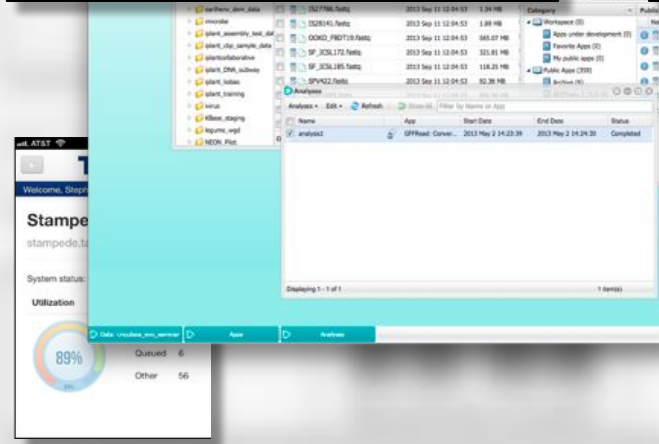
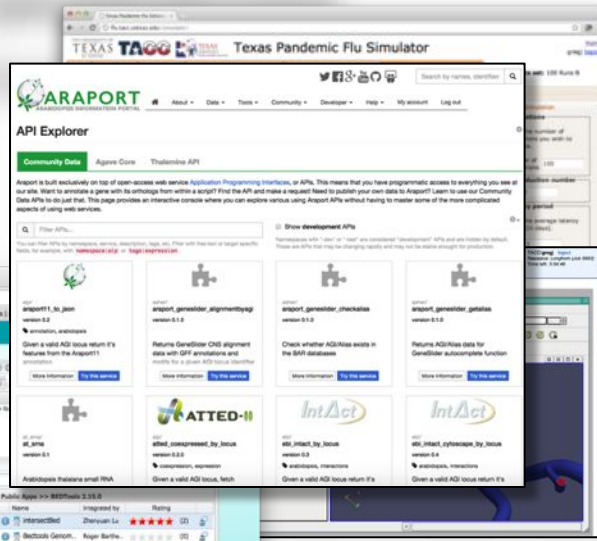
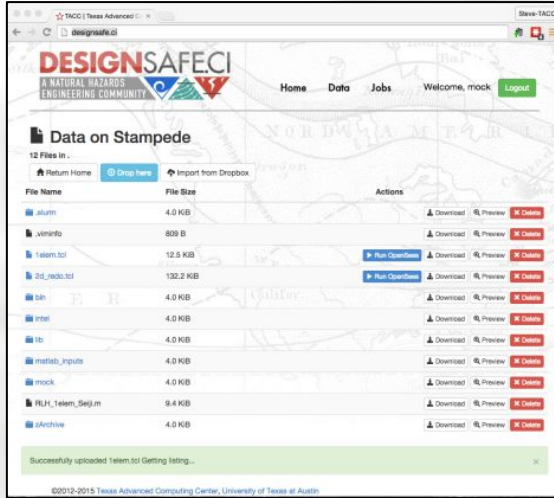






MANY DOMAIN SCIENTISTS ARE NOT EXPERTS AT COMPUTING TECHNOLOGY.  
CREATE PURPOSE-BUILT, HIGHLY INTUITIVE INTERFACES





## Point-and-click interfaces

- Data management, sharing, and analysis
- Publishing reproducible analysis workflows
- Discovery of new or updated tools and data
- Interactive visualization of results

Backed by world-class computing and data capacity



The image shows two overlapping windows. The background window is RStudio, displaying the R console with version 3.0.3 (2014-03-06) and a 'Warm Puppy' theme. The foreground window is a Jupyter notebook titled 'pyspark\_genome\_example' running on a JupyterHub instance. The notebook content includes an introduction to the pyspark library for genomics, a note about using python3, and a code cell with the following imports and initialization:

```
In [1]: import string, os
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline
from IPython.display import Image, display, Math, Latex, SVG, HTML
import numpy as np
from scipy.cluster.hierarchy import linkage, dendrogram
from scipy.spatial.distance import pdist
# from urllib2 import urlopen
from urllib.request import urlopen
import pyspark
sc = pyspark.SparkContext('local[*]')
```

Below the code cell, the word 'Pathogens' is visible, indicating the start of the next section in the notebook.

## Hosted SaaS

- JupyterHub notebooks
- Rstudio
- Web-based VNC

Also, backed by  
world-class computing  
and data capacity



## Getting Started



### Launch New Instance

Browse Atmosphere's list of available images and select one to launch a new instance.



### Browse Help Resources

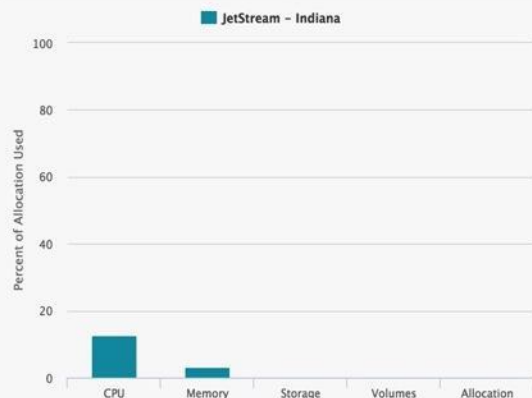
View a video tutorial, read the how-to guides, or email the Atmosphere support team.



### Change Your Settings

Modify your account settings, view your resource quota, or request more resources.

## Resources in Use [Need more?](#)



1 Instances

active - deploy\_error



0 Volumes

## Community Activity

**edwintest3** created an image  
Nov 16, 2015 02:31 am  
[MAKER-P 2.28 with CCTools 5](#)

**edwintest3** created an image  
Nov 16, 2015 02:31 am  
[TSW Workshop Williams 1.2](#)

**atmoadmin** created an image  
Oct 23, 2015 12:06 am  
[Trusty Tahr \(x64\)](#)

**atmoadmin** created an image  
Oct 23, 2015 12:06 am  
[cirros-0.3.4-x86\\_64](#)

**atmoadmin** created an image  
Oct 23, 2015 12:06 am  
[CentOS-7-x86\\_64-GenericCloud-20150628\\_01](#)

**atmoadmin** created an image  
Oct 23, 2015 12:06 am  
[CentOS-6-x86\\_64-GenericCloud-1508](#)

**atmoadmin** created an image  
Oct 23, 2015 12:06 am  
[CentOS-7-x86\\_64-GenericCloud-1508](#)

## Instance History (5 instances launched)



Updated a few seconds ago

## Easy to use Cloud Computing

- Atmosphere (Cyverse)
- Jetstream (IU,UA,TACC)
- Chameleon (UC,TACC)

Cloud consoles are aimed at sysadmins and unintuitive.

We're changing that with improved UX and support

- APIs are still available
- No cost to end user



GIVE EXPERTS ACCESS TO EVERY SINGLE ONE OF YOUR BUILDING BLOCKS.  
WEB SERVICE APIs EVERYWHERE. AUGMENT WITH PROFESSIONAL TOOLING.



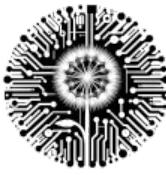


# TECHNOLOGY

- Cloud Computing
- Python
- JSON
- Docker
- Redis
- AND...







# EUREKA!





# POLLS

