



ADMI 2024

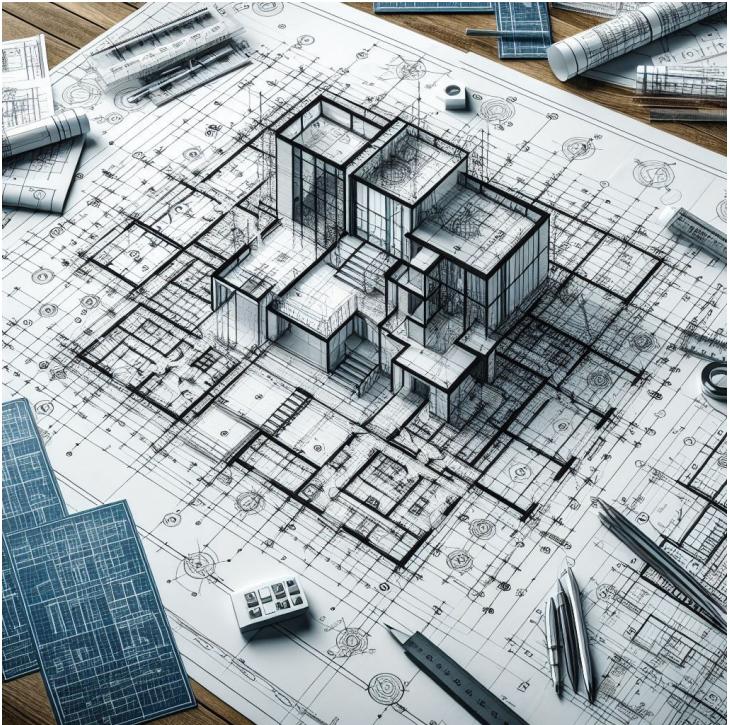
Code-a-thon



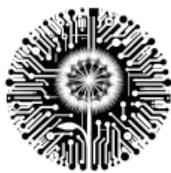


THE PLAN

- 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!

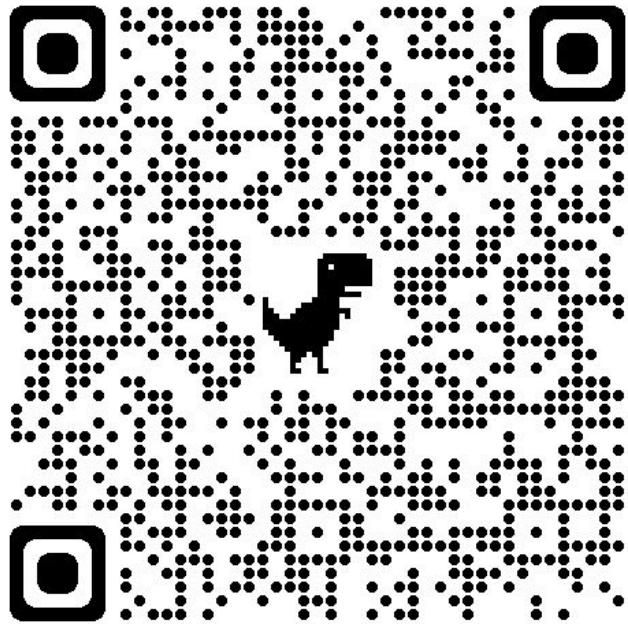


SGX3
Extend. Expand. Exemplify.



BUT FIRST!

It's time to Jam.

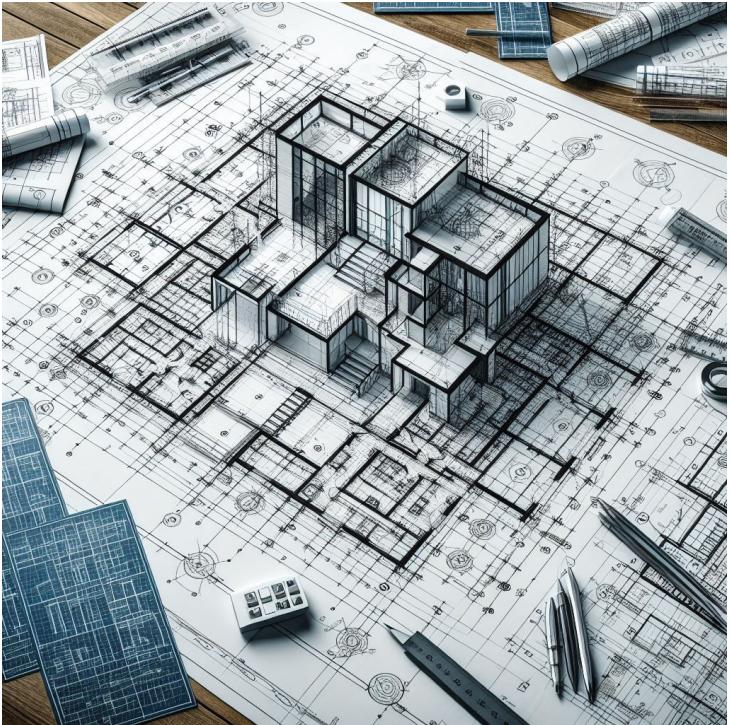


SGX3
Extend. Expand. Exemplify.

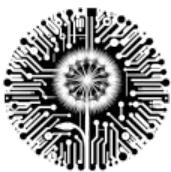


THE PLAN

- 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!



SGX3
Extend. Expand. Exemplify.



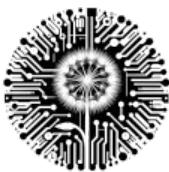
WHAT ARE WE DOING HERE?

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.**



SGX3
Extend. Expand. Exemplify.

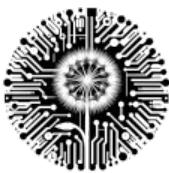


THE CODE-A-THON!

Code-a-thons have a more narrow **focus** primarily on **iterative coding, algorithmic development, over-the-shoulder peer coding**



SGX3
Extend. Expand. Exemplify.



THE CODE-A-THON!

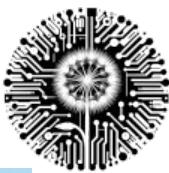
Participants are going to engage in coding challenges or competitions, where each challenge builds on the previous challenge.

These challenges are algorithmic or data structure-related and each challenge combines together to become a major project.



SGX3
Extend. Expand. Exemplify.

SOFTWARE ENGINEERING?

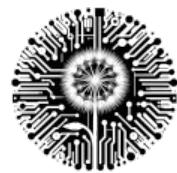


- What is Software Engineering
 - Requirements Gathering
 - Software Architecture
 - Coding and Programming
 - Software Testing and Debugging
 - Software Maintenance
 - Software Project Management
 - Software Quality
 - Software Metrics
 - Software Development Models & Architecture

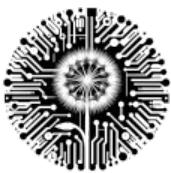


SGX3
Extend. Expand. Exemplify.

SOFTWARE ENGINEERING vs SCIENCE GATEWAYS vs HIGH PERFORMANCE COMPUTING



SOFTWARE ENGINEERING AND SCIENCE GATEWAYS



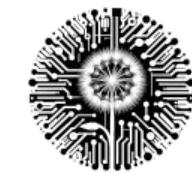
- What is a Science Gateway?

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



SGX3
Extend. Expand. Exemplify.

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**



SGX3
Extend. Expand. Exemplify.

"COMPUTING FOR THE ENDLESS FRONTIER"

Dan Stanzione

Executive Director, Texas Advanced Computing Center

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

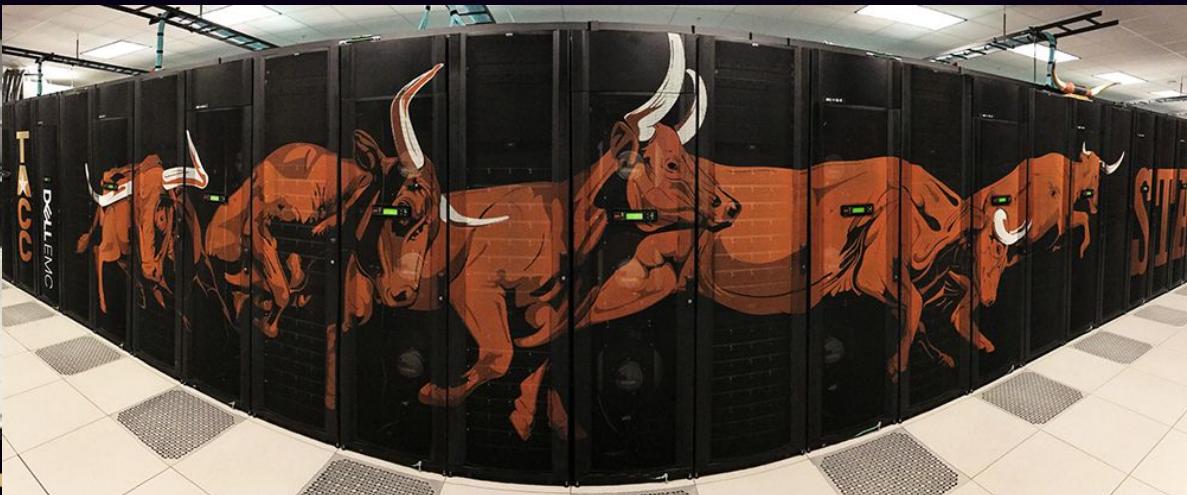
RAPID GROWTH FROM THEN TO NOW...

- ▶ 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)

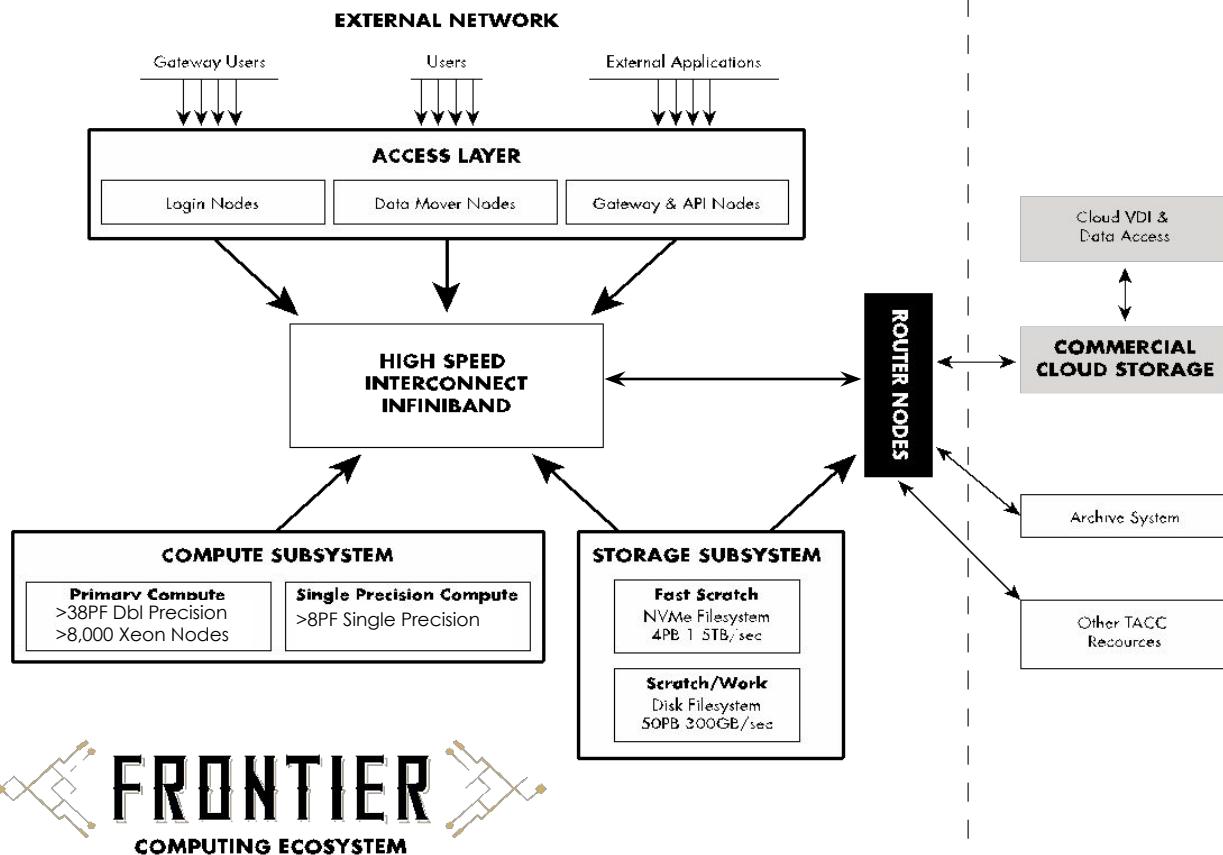


FRONTERA SYSTEM --- HARDWARE

- ▶ 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

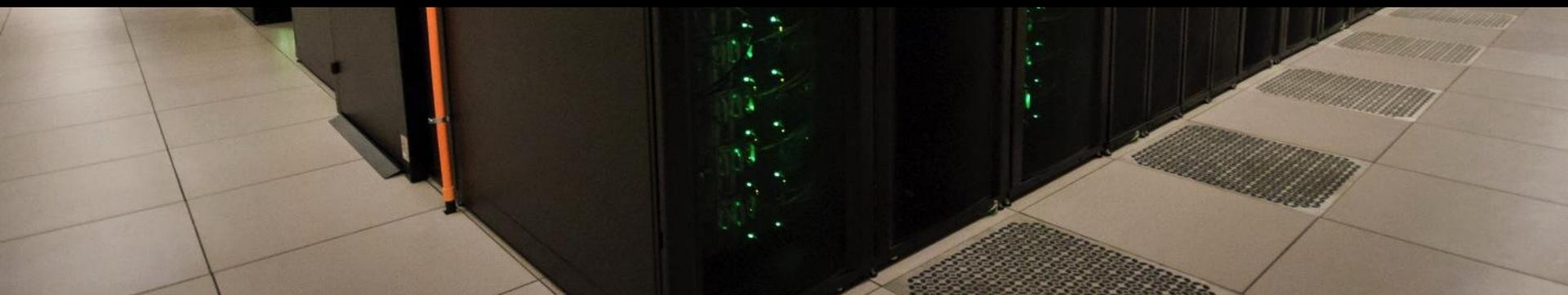
PICTUREQUOTES . com

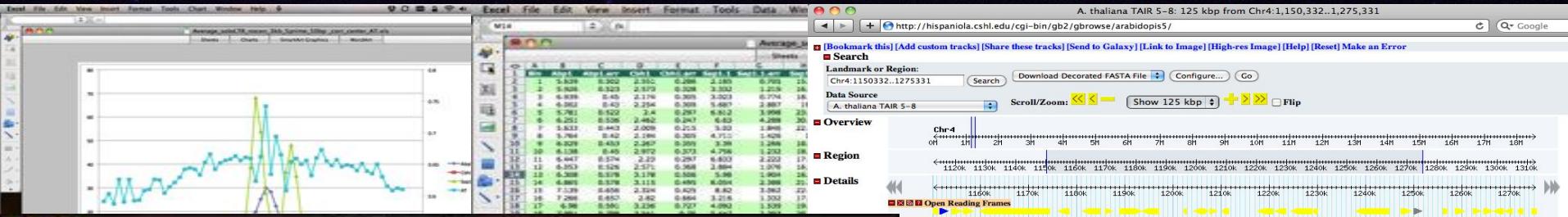
HOW DO WE HELP RESEARCHERS WITH SUCH
DIVERSE NEEDS AND BACKGROUNDS?





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





```
gcc -c -g -Wall -O2 -D_FILE_OFFSET_BITS=64 -D_USE_KNETFILE -D_CURSES_LIB=1 knetfile.c -o knetfile.o
gcc -c -g -Wall -O2 -D_FILE_OFFSET_BITS=64 -D_USE_KNETFILE -D_CURSES_LIB=1 bam_sort.c -o bam_sort.o
gcc -c -g -Wall -O2 -D_FILE_OFFSET_BITS=64 -D_USE_KNETFILE -D_CURSES_LIB=1 sam_header.c -o sam_header.o
```

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



The collage consists of six screenshots:

- DESIGNSAFE.CI**: A dashboard showing a map of the United States with hazard data overlays. It includes tabs for Home, Data, Jobs, Welcome, mock, and Logout.
- Data on Stampede**: A file management interface showing a list of files (e.g., alarm, telem.tcl, 2d_redo.tcl) with download, preview, and delete actions.
- Texas Pandemic Flu Simulator**: A web-based simulation tool with a header for TEXAS TACO TEAM and a search bar for names, identifier.
- API Explorer**: A service-oriented architecture (SOA) tool for managing APIs, showing a list of available services like 'aport_to_json' and 'aport_generator_alignmentgap'.
- Stampede**: A monitoring interface for the Stampede supercomputer, showing utilization (89%), queued jobs (6), and other system status.
- VWDJ SERVER**: A project management interface showing a list of projects and their details, such as 'Job Output Project' and 'example_01.fasta'.

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 screenshot shows a dual-pane interface. On the left, an RStudio session is running on port 51114, displaying the R console output for version 3.0.3. On the right, a Jupyter notebook titled "pyspark_genome_example" is open in a browser window. The notebook contains a section titled "A Genomics Example Using the pyspark Library" with a descriptive text block and a code cell labeled "In [1]". The code cell contains Python 3 code for importing various libraries and initializing a SparkContext.

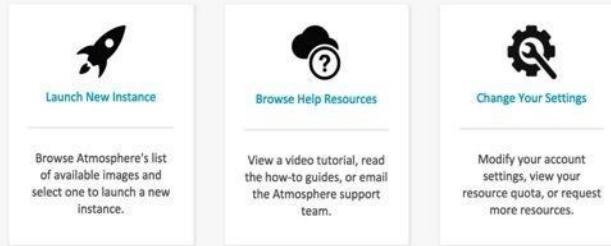
```
In [1]: import string, os
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline
from IPython.display import Image, display, Math, Latex, SVG, HTML
import numpy as np
from scipy.cluster.hierarchy import linkage,dendrogram
from scipy.spatial.distance import pdist
# from urllib2 import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

Hosted SaaS

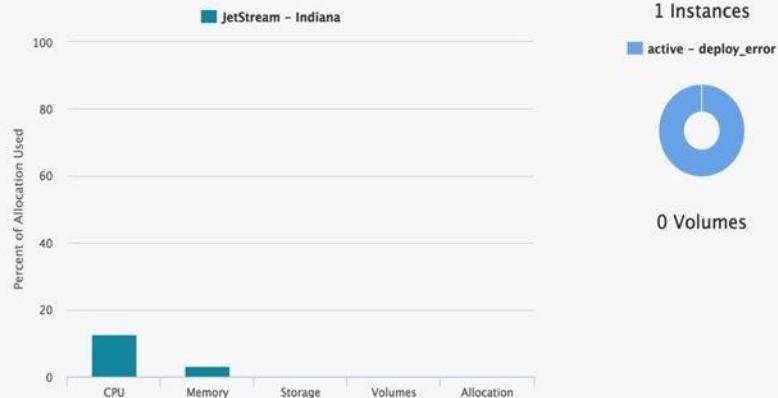
- JupyterHub notebooks
- Rstudio
- Web-based VNC

Also, backed by world-class computing and data capacity

Getting Started



Resources in Use Need more?



Instance History (5 instances launched)



Updated a few seconds ago

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
- atmodadmin created an image Oct 23, 2015 12:06 am Trusty Tahr (x64)
- atmodadmin created an image Oct 23, 2015 12:06 am cirros-0.3.4-x86_64
- atmodadmin created an image Oct 23, 2015 12:06 am CentOS-7-x86_64-GenericCloud-20150628_01
- atmodadmin created an image Oct 23, 2015 12:06 am CentOS-6-x86_64-GenericCloud-1508
- atmodadmin created an image Oct 23, 2015 12:06 am CentOS-7-x86_64-GenericCloud-1508

Easy to use Cloud Computing

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

Cloud consoles are aimed at sysadmins and unintuitive.

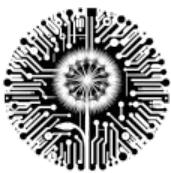
We're changing that with improved UX and support

- 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.



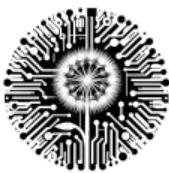


TECHNOLOGY

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



SGX3
Extend. Expand. Exemplify.

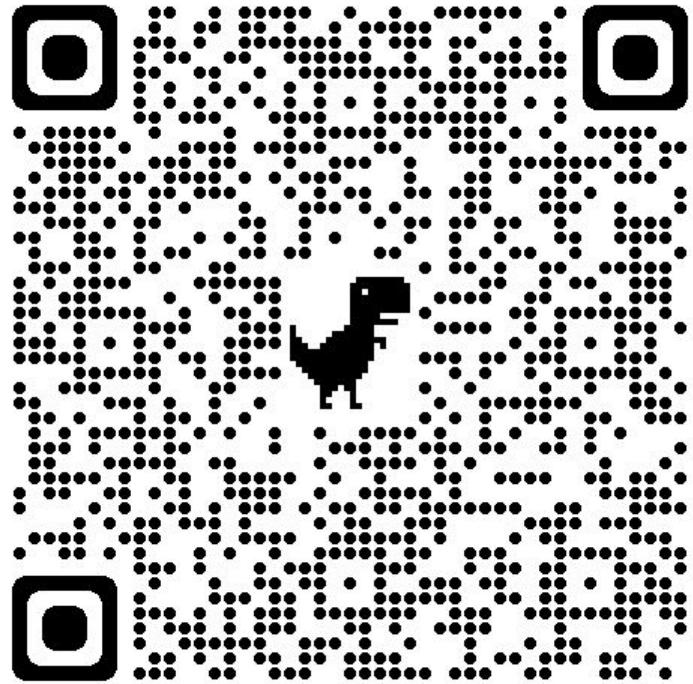
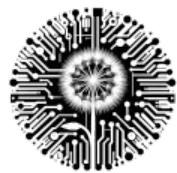


EUREKA!



SGX3
Extend. Expand. Exemplify.

POLLS



SGX3
Extend. Expand. Exemplify.