EXPANDING NRP USAGE WITH JUPYTER AT CAL STATE SAN BERNARDINO

YOUNGSU KIM ASST PROF OF MATH & HPC FACULTY FELLOW
DUNG VU HPC ANALYST

5th NRP
3/22/2024

ACKNOWLEDGEMENT

Part of our presentation was borrowed and modified from the work of our CFO & VP of FTO Sam Sudhakar

CALIFORNIA STATE UNIVERSITY SAN BERNARDINO (CSUSB)

- part of CSU system, 23 campuses
- located in near San Bernardino Mountains
- more than 70 traditional degree programs
- education credential and certificate programs, a doctoral program
- R2 in 2022 but primary focus on education

CSUSB (cont'd)

- 1,100 faculty members
- 19,000 students
- Serves 2 of CA's largest counties
- Hispanic Serving Institution
- 57% Pell Grant recipients
- Many student oriented projects

FACULTY LED RESEARCH

- CSUSB is in the top five CSU institutions for Faculty-Led Research
  $44M in federal funding in 2022-2023
- Most funded research projects did not require high-performance computing
- But when needed, faculty went on their own to garner resources from their colleagues and partner institutions

THE NEED

- Needed to find resources for faculty seeking HPC resources
- Didn’t necessarily have the funding, expertise, or the resources to build an infrastructure in-house
- Experimented with building an HPC Cluster with hopes of securing NSF funding to build an infrastructure

EXPLORATION OF RESOURCES OUTSIDE CSUSB

- In 2015 Library Dean Caballero connected our faculty with Extreme Science and Engineering Discovery Environment (XSEDE now ACCESS) team
- Transition from XSEDE to San Diego Supercomputing Center
- In 2017 CENIC introduced us to the Pacific (National) Research Platform

CSUSB JOINS THE NRP

- In June 2017 CSUSB became a part of the NRP
  the second CSU to become a part of this prestigious research network
- Created a Science DMZ to get connected to the Platform
- In 2018 James Macdonell and Dr. Dung Vu attended the FIONA workshop at the CENIC Annual Conference
- Currently CSUSB has three nodes dedicated to the NRP
  - perfsonar.csusb.edu
  - csusb-sci-103-11.csusb.edu
  - gpu-01.csusb.edu

HPC PROGRAM TEAM

- **Dr. Dung Vu** Information Security & Emerging Technologies Lead Research Computing Analyst
- **James Macdonell** Interim Information Security Officer
- **Dr. Youngsu Kim** 2021-2024 HPC Faculty Fellow & Assistant Professor of Mathematics
- **Gerard Au** Chief Information Officer
- **Dr. Bradford Owen** Associate Vice President of Faculty Development and Chief Academic Technology Officer

HPC ACTIVITIES AT CSUSB

Faculty-led Projects

- Dr. Cousins, Dr. Kim, Dr. Meyer, Dr. Pham, Dr. Ratnasingam, Dr. Salloum, Dr. Zhang, Dr. Alavi & Dr. Mohabbati, Dr. Becerra, Dr. Liszka & Dr. Kraemer

Research Groups

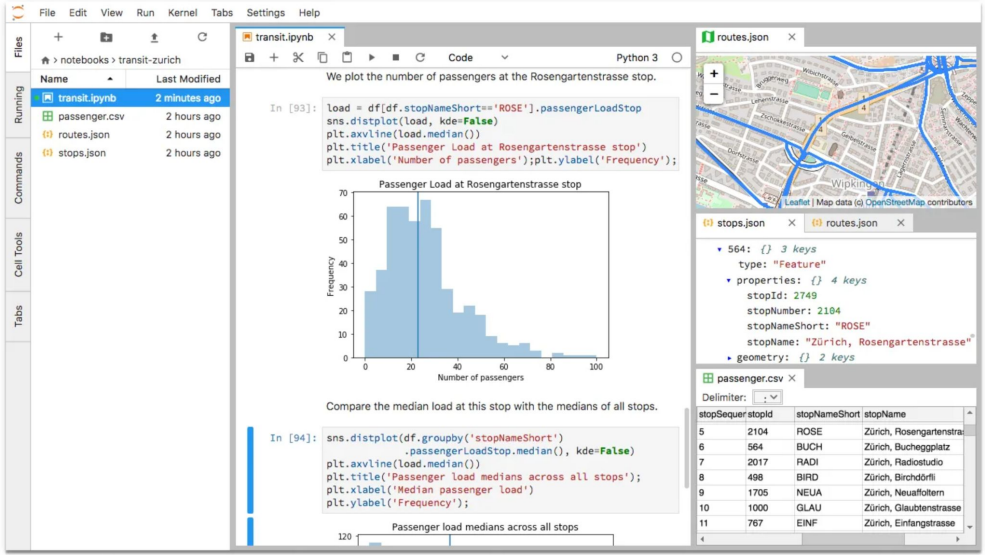
- Center for Advanced Functional Materials, Data Analytics Working Group, William and Barbara Leonard Transportation Center

We Support Courses

- BIOL 5050, CSE 5160, CHEM 5002, IST 2510, IST 6110, IST 6620, MATH 3465, MATH 6016

Nearly all the projects, workshops, and classes are supported by JupyterHub

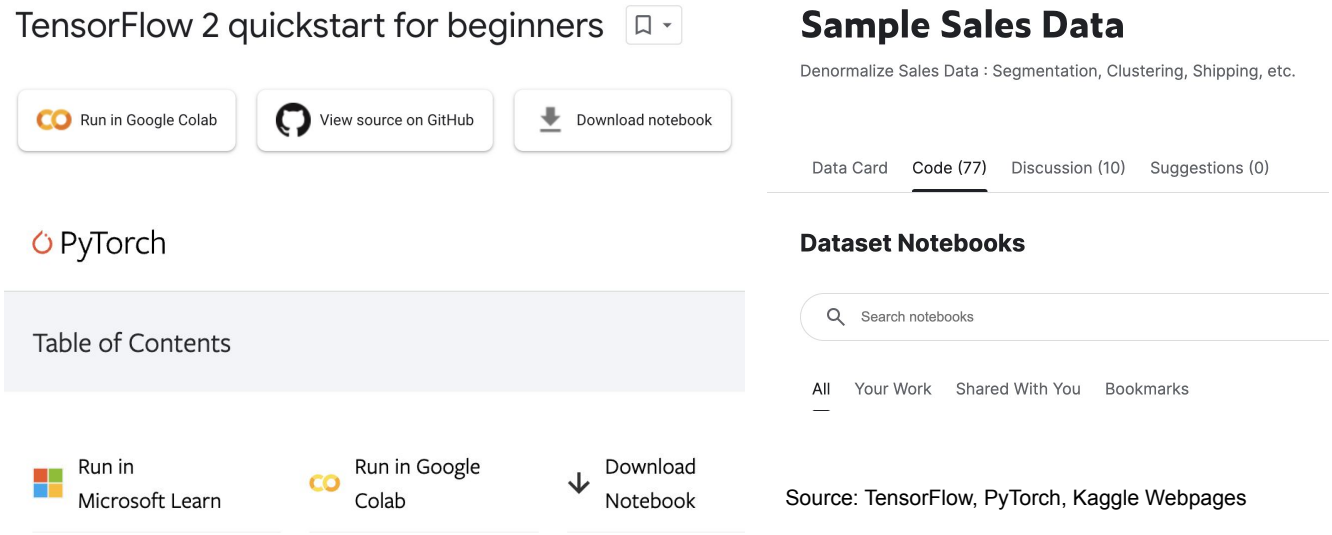
JupyterLab/Hub



JupyterLab is a highly extensible,
feature-rich notebook authoring application
and editing environment,
and is a part of Project Jupyter...

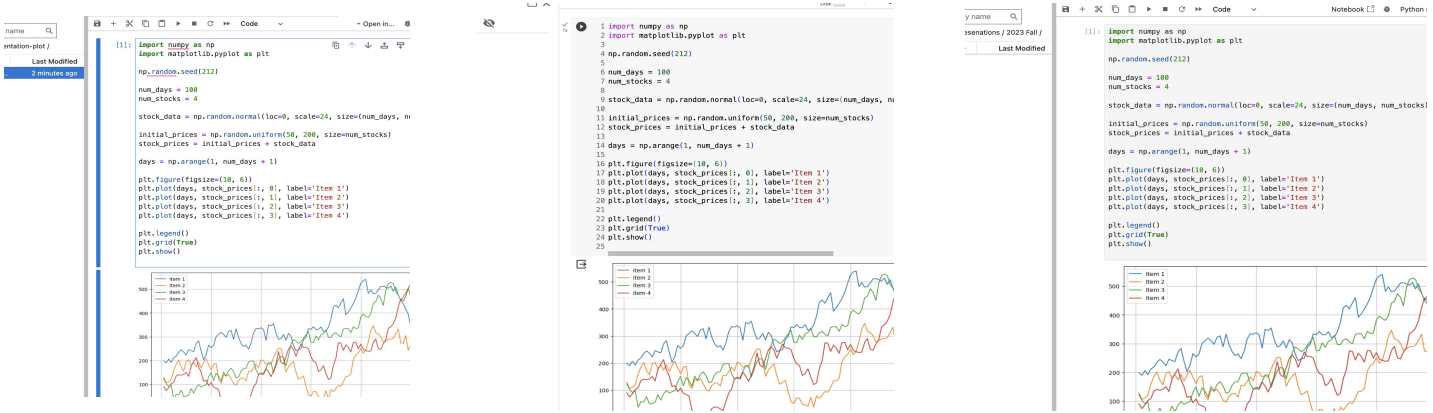
Source: JupyterLab Documentation

JupyterLab



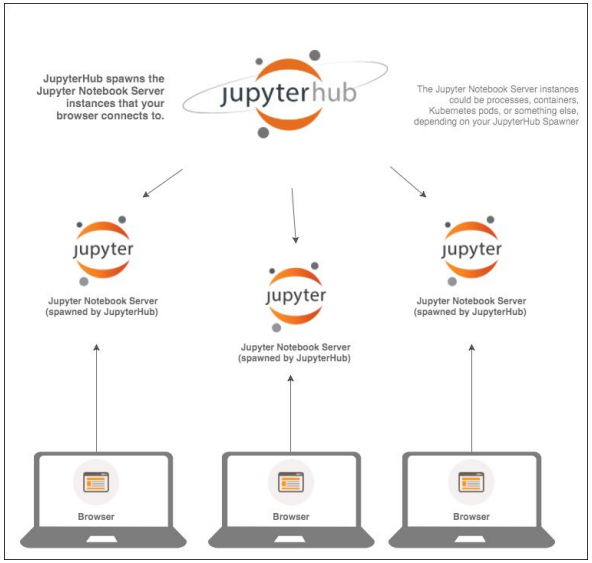
Source: TensorFlow, PyTorch, Kaggle Webpages

JupyterLab (cont'd)

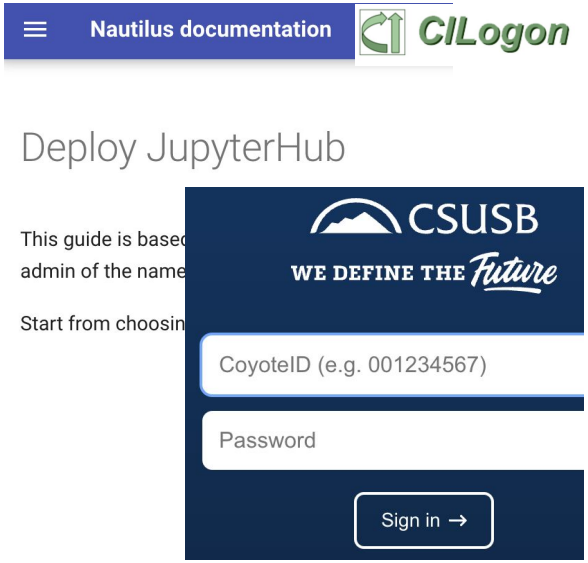


JupyterLab Notebook on Google Colab, My Office Computer, NRP server

JupyterHub/NRP

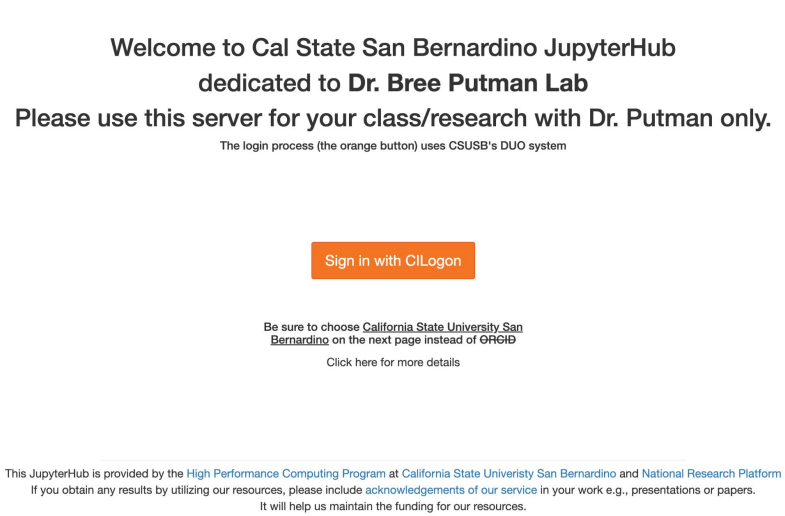


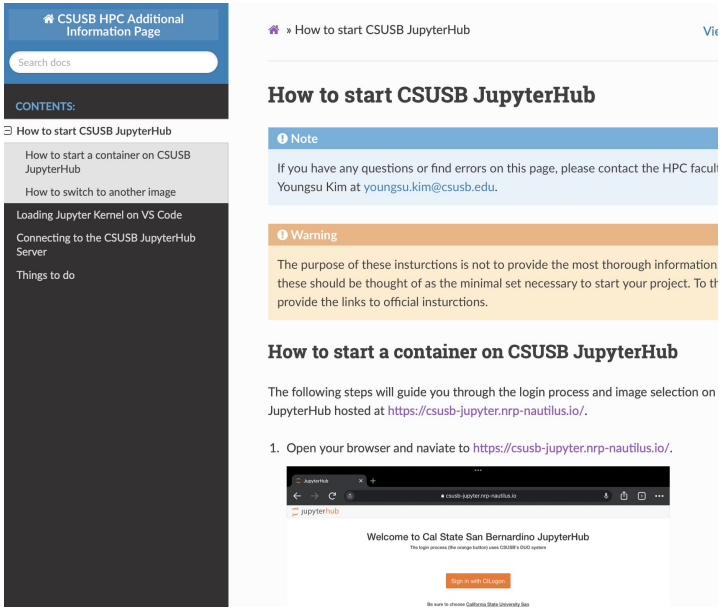
Source: OpenCredo.com



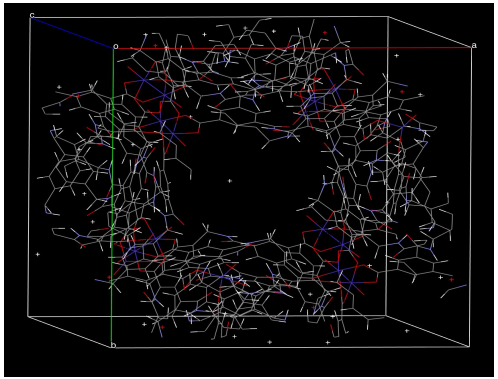
NRP Documentation, CILogon, CSUSB IDP

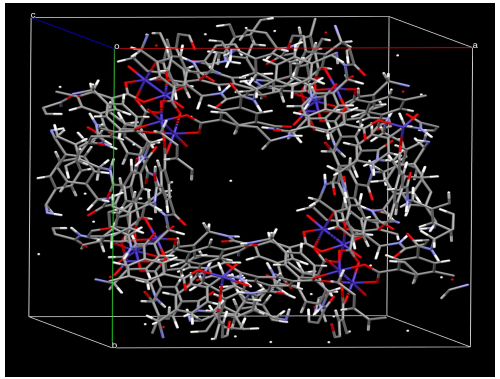
CUSTOMIZED PAGE & SUPPORT

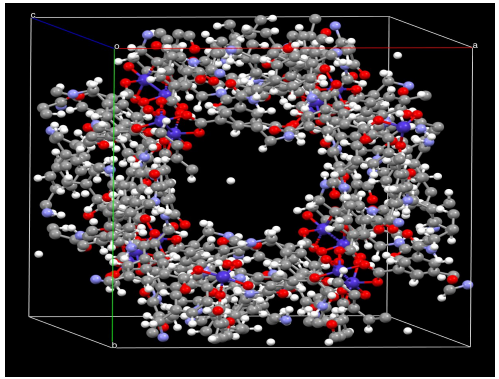




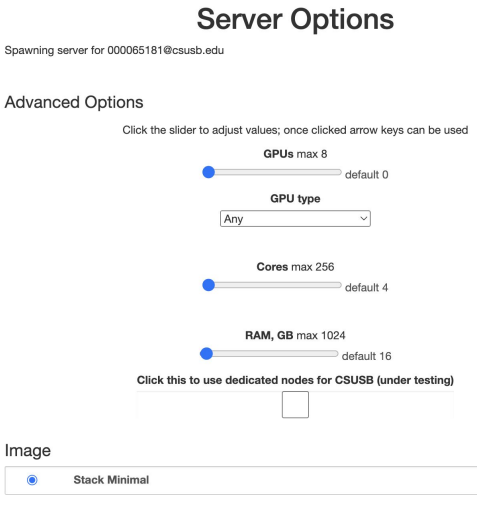
WORK OF DR COUSINS WITH VASP

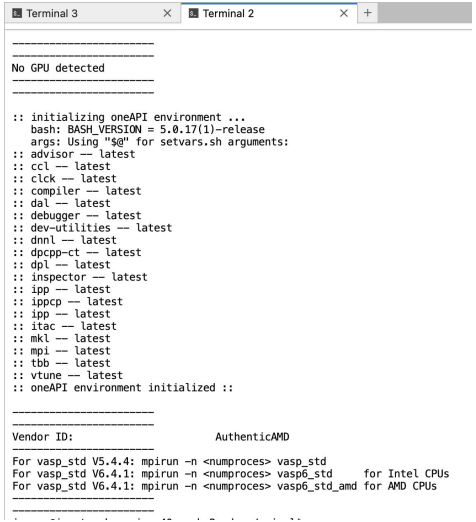






DR COUSINS (Cont'd)

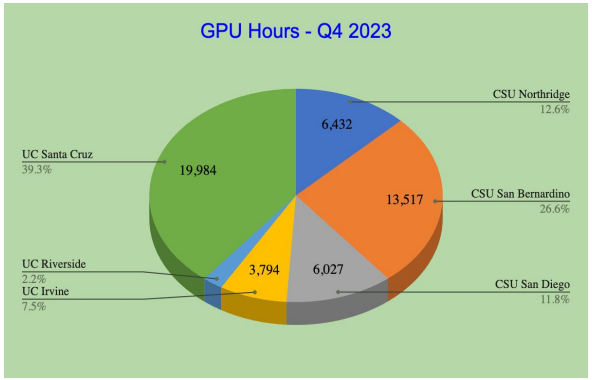




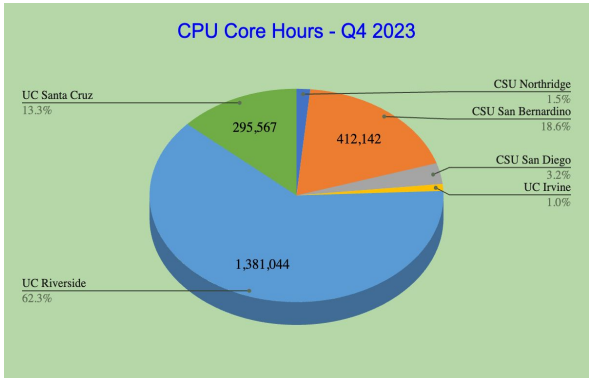
JupyterHub Dynamic Resource Allocation/Terminal Shell and NRP Resource Usage

CSUSB RESOURCE USAGE AMONG CALIFORNIA UNIVERSITIES

13.5k GPU hrs & 412k CPU hrs Q4 2023 (AWS Estimate of $50k-60k)



50,893 GPU Hrs

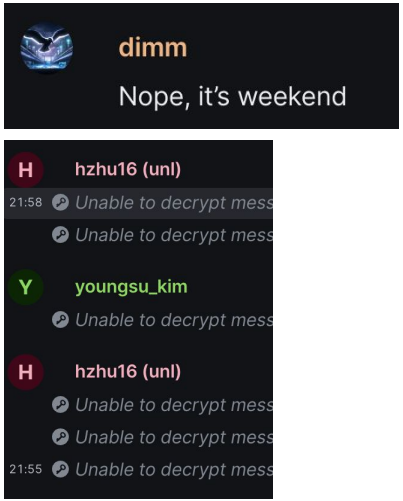


2,215,436 CPU Hrs

INTERMISSION

Thank the NRP team and special thanks to

- Tom DeFanti, Larry Smarr
- Dima, Huijun, Ashton, Mahidhar



How Jupyter Notebooks are Provided

Jupyter approach, by far, is the most efficient and scalable way so that everyone can access HPC powers at their fingertips.

- **Jupyterhub for CSUSB's members**: https://csusb-jupyter.nrp-nautilus.io/
  - over 450 distinct people who logged onto CSUSB's JupyterHub
  - Available to all CSUSB members, students, and other collaborative universities.
  - Leverage NRP's resources to deploy research computing packages.
  - Include customized software stacks.
  - Production readiness: 24x7 availability with dedicated tech support.
  - Some limitations to preserve resources.
-

How Jupyter Notebooks are Provided (Cont'd)

- **Jupyterhub dedicated for research**: https://csusb-hpc.nrp-nautilus.io/
  - For faculty / researchers
  - No limitation with very large HPC resources - high-end GPU.
  - Uninterruption - days, weeks until processing jobs are completed
  - Dedicated hardware and tech support

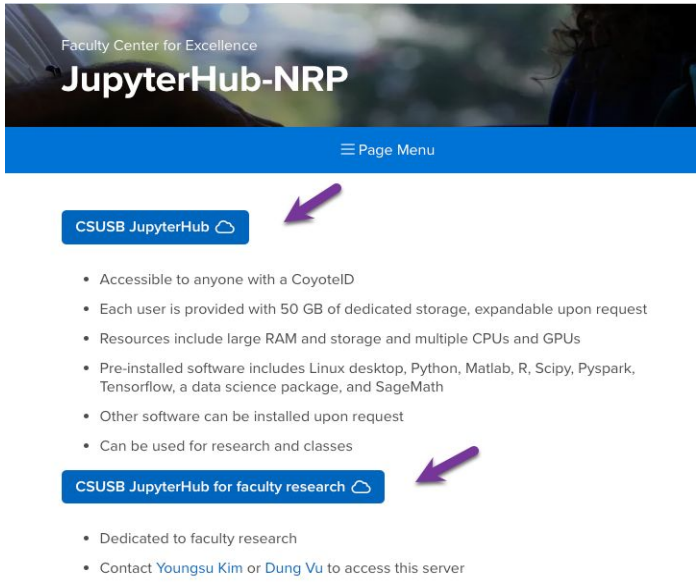
How Jupyter Notebooks are Provided (Cont'd)

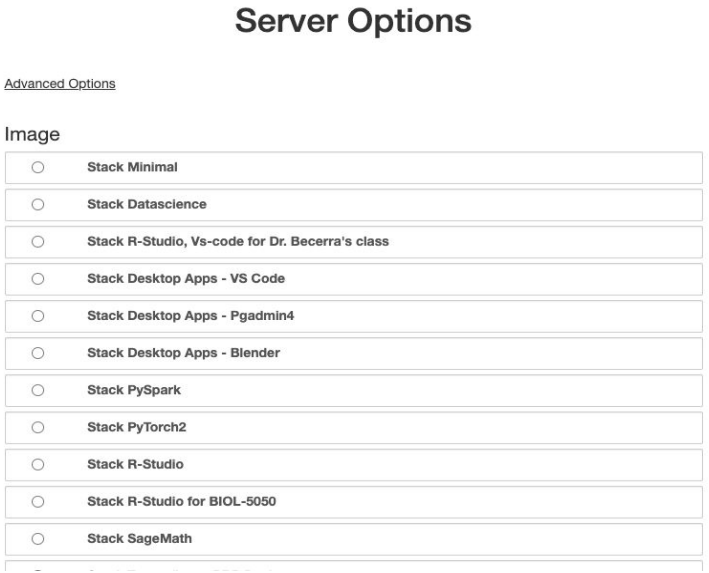
- **Multiple dedicated Jupyter Hubs and dedicated systems**:
  - Dedicated to research group / projects
  - Customized/dedicated software/stacks (VASP, blender, etc..)
  - Dedicated / reserved HPC resources / storages
  - Restricted access
  - Secured shared data.
  - PI administration

How Jupyter Notebooks are Provided (Cont'd)

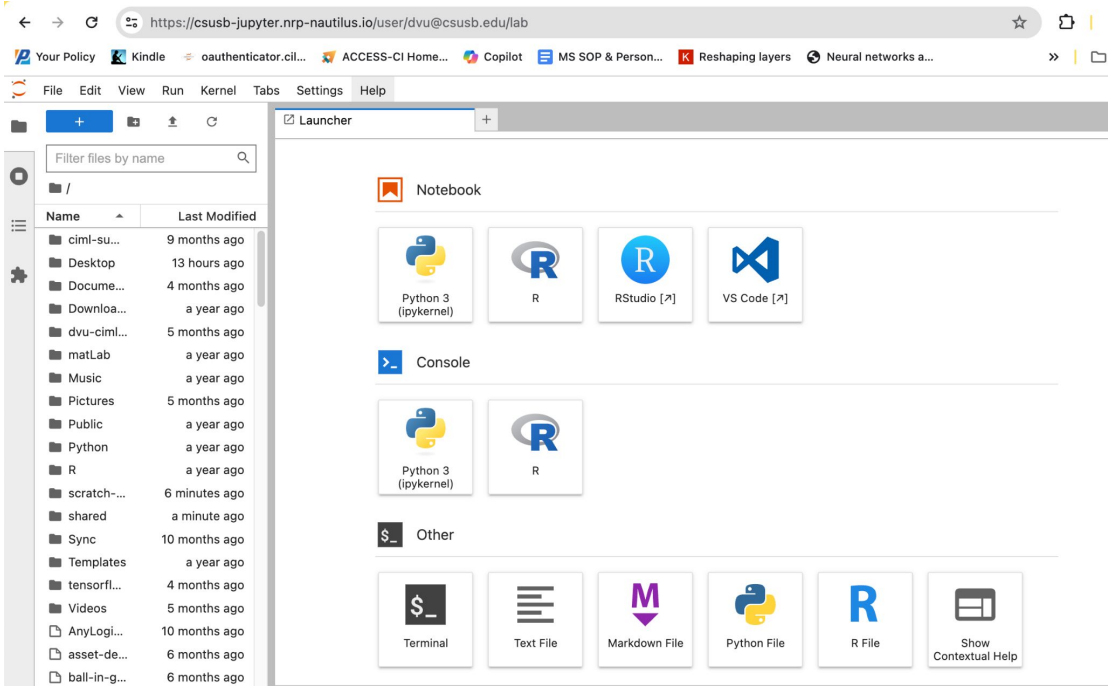
- **Current dedicated Jupyter Hubs and dedicated systems:**
  - https://csusb-metashape.nrp-nautilus.io: 3D modeling
  - https://csusb-vasp1.nrp-nautilus.io Viena Ab initio Simulation package (VASP)
  - https://csusb-cousins-lab.nrp-nautilus.io: VASP simulation
  - https://csusb-becerra.nrp-nautilus.io AI/ML project
  - https://csusb-biol-5050.nrp-nautilus.io: Biology course
  - https://csusb-cse-salloum.nrp-nautilus.io Summer Research
  - https://csusb-drhamoudahub.nrp-nautilus.io Data Analytics
  - https://csusb-ratnasingam.nrp-nautilus.io Data Analytics
  - https://csusb-zhang.nrp-nautilus.io AI/ML project

How Jupyter Notebooks are Provided (Cont'd)

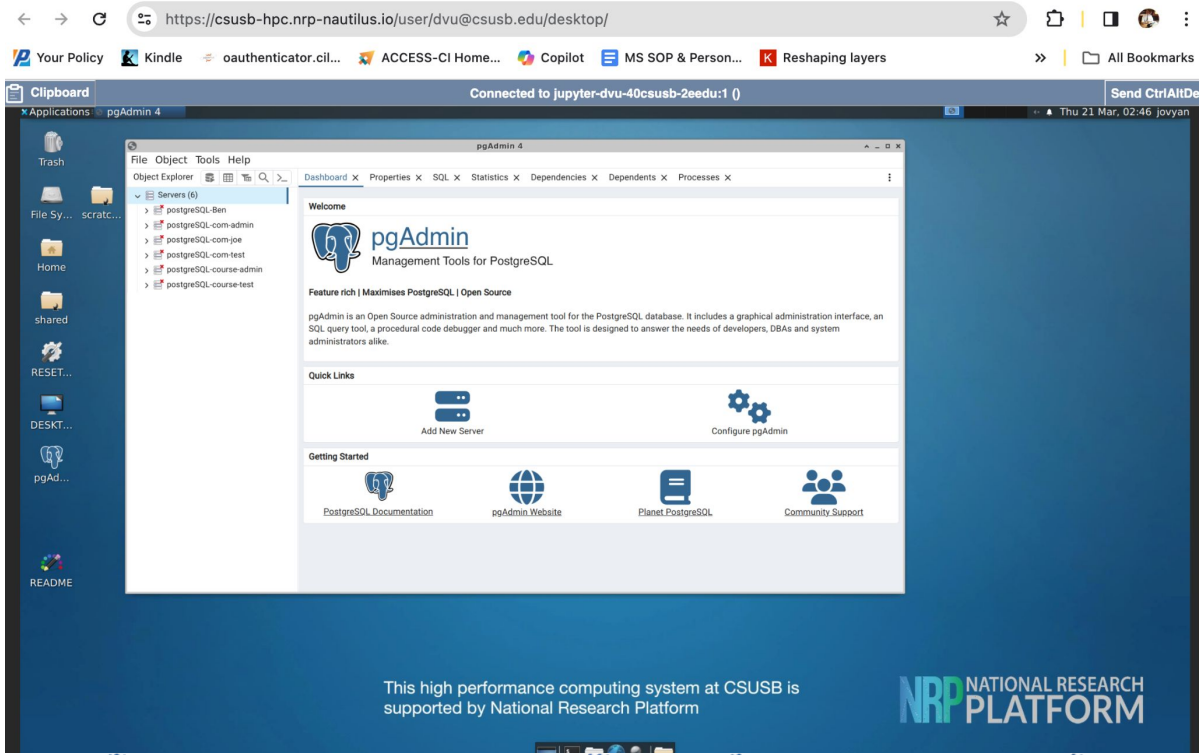




A Typical JupyterLab



A Typical HPC Desktop



Monitoring Running Pods

```
+ kubectl get pods -n prp-dvu-csusb -o wide
NAME                              READY   STATUS    RESTARTS   AGE    IP              NODE                              NOMINATED NODE   READINESS GATES
csusb-metashape-bc54586c-jcv4r    1/1     Running   0          6d2h   10.244.32.252   ry-gpu-14.sdsc.optiputer.net      <none>           <none>
+ kubectl get pod -n csusb-hpc -o wide
NAME                                                 READY   STATUS    RESTARTS      AGE     IP              NODE
hub-5dfbd58c4b-rr9f8                                 1/1     Running   0             5d18h   10.244.243.174  gpn-fiona-mizzou-8.rnet.missouri.edu
jupyter-aswin-2enalluri-40csusb-2eedu                1/1     Running   0             5d17h   10.244.16.234   ry-gpu-13.sdsc.optiputer.net
jupyter-david-2evargasmonroy0554-40coyote-2ecsusb-2eedu  1/1     Running   0             4d9h    10.244.91.8     gpn-fiona-mizzou-6.rnet.missouri.edu
jupyter-suthakaran-2eratnasingam-40csusb-2eedu       1/1     Running   0             4d15h   10.244.133.177  gpn-fiona-mizzou-7.rnet.missouri.edu
postgres-ben-749fc57dd8-dfz8z                        1/1     Running   1 (39d ago)   39d     10.244.158.241  hcc-nrp-shor-c6025.unl.edu
proxy-7f9d5644c9-x752b                               1/1     Running   0             11d     10.244.243.148  gpn-fiona-mizzou-8.rnet.missouri.edu
syncthing-csusb-5956d44fbf-rx5jb                     0/1     Pending   0             8d      <none>          <none>
+ kubectl get pod -n csusb-cousins-lab -o wide
NAME                              READY   STATUS    RESTARTS        AGE   IP              NODE                      NOMINATED NODE   READINESS GATES
hub-6b769d468-vkz4t               1/1     Running   1 (6d14h ago)   12d   10.244.237.198  suncave-12                <none>           <none>
jupyter-kcousins-40csusb-2eedu    1/1     Running   0               17h   10.244.57.67    node-2-6.sdsc.optiputer.net  <none>           <none>
proxy-7c944fb594-4dbdp            1/1     Running   0               25d   10.244.19.31    fiona.nwsc.ucar.edu       <none>           <none>
+ kubectl get pod -n csusb-mpi -o wide
NAME                                                 READY   STATUS    RESTARTS   AGE     IP              NODE                              NOMI
hub-669d7f4546-mkvw8                                 1/1     Running   0          4d5h    10.244.119.102  suncave-5                         <non
jupyter-david-2eilie2542-40coyote-2ecsusb-2eedu       1/1     Running   0          7d12h   10.244.241.248  ry-gpu-12.sdsc.optiputer.net      <non
jupyter-reymundo-2ealatorre8223-40coyote-2ecsusb-2eedu  1/1     Running   0          10d     10.244.50.221   rci-nrp-cpu-04.sdsu.edu           <non
proxy-7cf448fc6c-d5hhn                               1/1     Running   0          11d     10.244.243.162  gpn-fiona-mizzou-8.rnet.missouri.edu  <non
+ kubectl get pod -n csusb-khan -o wide
NAME                                      READY   STATUS    RESTARTS   AGE     IP              NODE                              NOMINATED
hub-6485f8f697-gzmwt                      1/1     Running   0          4d13h   10.244.56.115   node-2-11.sdsc.optiputer.net      <none>
jupyter-nathan-2ekelly0642-40coyote-2ecsusb-2eedu  1/1     Running   0          35h     10.244.26.101   ry-gpu-01.sdsc.optiputer.net      <none>
proxy-5c5944d7f9-5r89q                    1/1     Running   0          14d     10.244.142.101  gpn-fiona-mizzou-9.rnet.missouri.edu  <none>
(base) 000065181@IIS-ISFI-65681 ~ %
```

Current AI/ML - Data Science Courses

1. CSE 5160 Machine Learning
2. IST 2510 Big Data Analytics
3. BIOL 5050 - Biostatistics and Experimental Design
4. IST 6110 - Foundations of Analytics and Big Data
5. IST 6620 - Business Analytics and Decision-Making
6. MATH 3465 - Computational Statistics

Current AI/ML Projects

There are 14 ongoing AI / ML projects

1. **Dr. Becerra:**
   - Sleep Apnea
2. **Dr. Ronald Salloum**
   - Cloud Removal from Satellite Imagery, Dr. Ronald Salloum
3. **Dr. Yan Zhan**
   - Brain Tumor Classification with Deep Learning
   - Hybrid Recommendation with Machine Learning
   - Public Opinion Sentiment Analysis
   - Medical Notes Labelling with Machine Learning
   - Drug-Drug Interaction Prediction Based on Drug Feature Extraction and Drug Similarity

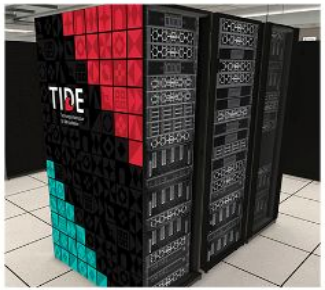
Current AI/ML Projects

4. **Dr. Bilal Khan**
    - Automated identification of sheep activity via hybrid deep learning and computer vision approaches
    - Real-Time Gun Detection in Video Streams using YOLOv8
    - A Graph Attention Network Based Approach for Interpretable and Domain-aware Modeling of a Water Treatment System
    - Exploring Wearable Interactive Technologies for Understanding and Assessing Autism Spectrum Disorder through Machine Learning Analysis of Communication and Social Interaction
5. **Dr. Jennifer Jin**
    - Personalized Health Assistant with Reinforcement Learning
    - GPT-Empowered Personalized eLearning System for Programming Languages
    - Classification of Remote Sensing Image data Using RSSCN-7 dataset

Contributions To NRP's Infrastructure

Hardware Details

Node Type	Qty.	Details
GPU	17	Dell PowerEdge R760XA (2x) Intel Xeon Silver 4410Y 2G CPU, 12C/24T (4x) Nvidia L40 GPU, 48 GB RAM 512 GB System RAM
GPU Advanced	1	Dell PowerEdge R750XA (2x) Intel Xeon Gold 6338 2G CPU, 32C/64T (4x) Nvidia A100 GPU, 80 GB RAM 512 GB System RAM
CPU	6	Dell PowerEdge R760 (2x) Intel Xeon Gold 6430 2.1G CPU, 32C/64T 768 GB System RAM
Storage	3	Dell PowerEdge R760 (2x) Intel Xeon Gold 6442Y 2.6G CPU, 24C/48T 256 GB System RAM 240 TB RAW Capacity



616 CPU Cores | 73 GPUs | 14.592 TB RAM

U.S. National Science Foundation

National Science Foundation, Campus Cyberinfrastructure: Regional Computing, Award 2346701

https://tide.sdsu.edu/

Publications

- Kimberley Cousins, “Finding new function among known organic and organometallic crystalline materials” accepted for presentation at CINF and SciMix, American Chemical Society National Meeting, New Orleans, LA, March 17-21, 2024.
- Sivaprasad Ghanta, J. Pham, et al, Atomic site preference, electronic structures, and magnetic properties of γ-brass type pseudo-binary Mn2Zn11–Ni2Zn11 at high Mn-contents, Journal of Alloys and Compounds
- Kylee Young, J. Pham, Computation-Guided Synthesis of Alkali–Transition-Metal–Halides. CRESTII Summer 2022 Research Poster.
- Jaiung Jun, Youngsu Kim, and Matthew Pisano, On Picard groups and Jacobians of directed graphs, arxiv:2302.10327.
- Dr. K. Liszka, Wadi el-Hudi Expedition: 3D modeling and Puplications

  https://wadielhudi.com/3d-modeling-at-wadi-el-hudi/

  https://wadielhudi.com/publications/

Future Plan

- With access to rich HPC resources, especially with the recent $1M NSF funded CSU-TIDE Cyberinfrastructure, the following are CSUSB's future endeavors
  - Promote HPC awareness in faculty and expand HPC support to research for all academic domains.
  - Expand support for HPC-related courses, and growing research groups that involve more student participation
  - Collaborate with CSE faculty, SDSC, NRP to provide AI/ML support: As AI/ML has been integrating with all research aspects, they are indispensable tools for researchers for even the most complex research questions.
  - Expand collaborations with CSU sisters to share HPC services.

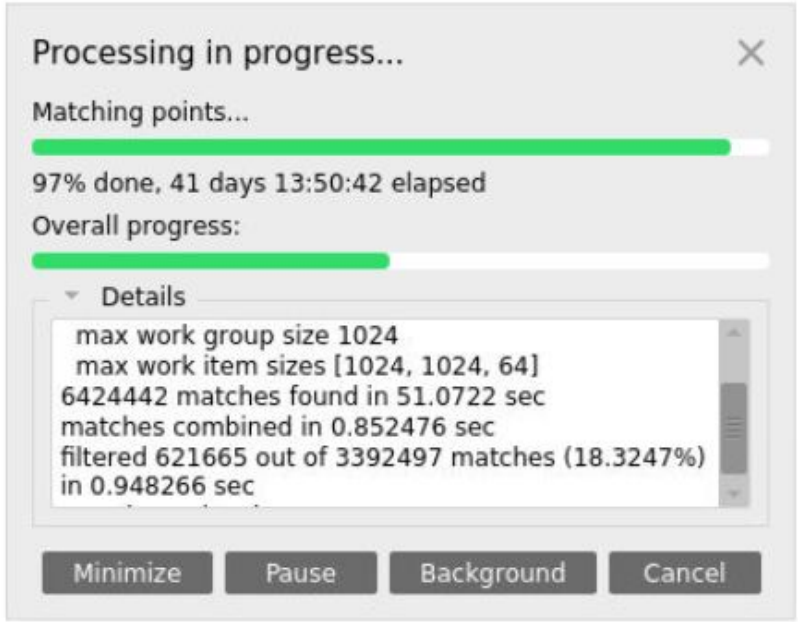
Acknowledgements

- NRP has not only provided computing resources, hands-on expertises, but also dedicated technical support and solutions
- It is a model to advance HPC programs, especially, for MSI who lack cyber infrastructure and HPC expertise
- CSUSB is able to provide research computing capability to its faculty, students as if we were a UC. Thanks NPR, especially
  - Dr. Smarr, Dr. Defanti, Mr. Graham, Dr. Mishin
- CENIC - Innovation Award:

CSUSB
WE DEFINE THE Future

Monitoring Running Pods

```
+ kubectl get pods -n prp-dvu-csusb -o wide
NAME                              READY   STATUS    RESTARTS   AGE    IP              NO
csusb-metashape-bc54586c-jcv4r    1/1     Running   0          6d2h   10.244.32.252   ry
+ kubectl get pod -n csusb-hpc -o wide
NAME                                                READY   STATUS    RESTARTS
hub-5dfbd58c4b-rr9f8                                1/1     Running   0
jupyter-aswin-2enalluri-40csusb-2eedu                1/1     Running   0
jupyter-david-2evargasmonroy0554-40coyote-2ecsusb-2eedu   1/1     Running   0
jupyter-suthakaran-2eratnasingam-40csusb-2eedu       1/1     Running   0
postgres-ben-749fc57dd8-dfz8z                       1/1     Running   1 (39d ago)
proxy-7f9d5644c9-x752b                              1/1     Running   0
syncthing-csusb-5956d44fbf-rx5jb                    0/1     Pending   0
+ kubectl get pod -n csusb-cousins-lab -o wide
NAME                              READY   STATUS    RESTARTS        AGE
hub-6b769d468-vkz4t               1/1     Running   1 (6d14h ago)   12d    10.244.237.198
jupyter-kcousins-40csusb-2eedu    1/1     Running   0               17h    10.244.57.67
proxy-7c944fb594-4dbdp            1/1     Running   0               25d    10.244.19.31
+ kubectl get pod -n csusb-mpi -o wide
NAME                                                READY   STATUS    RESTARTS   A
hub-669d7f4546-mkvw8                                1/1     Running   0          4
jupyter-david-2eilie2542-40coyote-2ecsusb-2eedu      1/1     Running   0          7
jupyter-reymundo-2ealatorre8223-40coyote-2ecsusb-2eedu   1/1     Running   0          1
proxy-7cf448fc6c-d5hhn                              1/1     Running   0          1
+ kubectl get pod -n csusb-khan -o wide
NAME                                          READY   STATUS    RESTARTS   AGE
hub-6485f8f697-gzmwt                          1/1     Running   0          4d13
jupyter-nathan-2ekelly0642-40coyote-2ecsusb-2eedu   1/1     Running   0          35h
proxy-5c5944d7f9-5r89q                        1/1     Running   0          14d
(base) 000065181@IIS-ISFI-65681 ~ %
```




```
10.244.26.101    ry-gpu-01.sdsc.optiputer.net            <none>
10.244.142.101   gpn-fiona-mizzou-9.rnet.missouri.edu    <none>
```

How HPC Resources Are Provided

- We provide full HPC assistance
  - Promote HPC resources with hardware and software availability
  - Approach faculty to learn about HPC needs and offer HPC support
  - Deploy software and build scalable and optimized HPC systems
  - Ensure resource availability, reliability, and stability
  - Collaborate with NRP tech to provide solutions
  - Update users with advanced and cutting-edge HPC features
  - We take responsibility to ensure faculty successfully complete their research – giving up is not a solution