

# ACES: Intermediate Python Programming In JupyterLab

Accelerating Workflows on a Composable Cyberinfrastructure

Zhenhua He

07/14/2024



# ACES TechLab

## **Lab I. JupyterLab (30 mins)**

We will load required modules and activate virtual environment and run JupyterLab on HPRC ACES portal.

## **Lab II. Data Exploration with Python(30 mins)**

We will go through some examples with a popular Python library Pandas for data exploration.

## **Lab III. Machine/Deep Learning (30 minutes)**

We will learn how to use PyTorch to build and train a simple image classification model with deep neural network (DNN).

# Lab I. JupyterLab



File Edit View Run Kernel Tabs Settings Help

Files

- + notebooks
- Name Last Modified
- Data.ipynb an hour ago
- Fasta.ipynb a day ago
- Julia.ipynb a day ago
- Lorenz.ipynb seconds ago**
- R.ipynb a day ago
- iris.csv a day ago
- lightning.json 9 days ago
- lorenz.py 3 minutes ago

Running

Commands

Cell Tools

Output View

lorenz.py

In this Notebook we explore the Lorenz system of differential equations:

$$\begin{aligned}\dot{x} &= \sigma(y - x) \\ \dot{y} &= \rho x - y - xz \\ \dot{z} &= -\beta z + xy\end{aligned}$$

Let's call the function once to view the solutions. For this set of parameters, we see the trajectories swirling around two points, called attractors.

```
In [4]: from lorenz import solve_lorenz
t, x_t = solve_lorenz(N=10)
```

sigma 10.00  
beta 2.67  
rho 28.00

A 3D plot of the Lorenz attractor, showing a complex, swirling trajectory in a 3D space. The plot is rendered with a green and blue color scheme, and the axes are labeled x, y, and z.

def solve\_lorenz(N=10, max\_time=4.0, sigma=10.0, beta=8./3, rho=28.0):  
 """Plot a solution to the Lorenz differential equations."""  
 fig = plt.figure()  
 ax = fig.add\_axes([0, 0, 1, 1], projection='3d')  
 ax.axis('off')  
  
 # prepare the axes limits  
 ax.set\_xlim((-25, 25))  
 ax.set\_ylim((-35, 35))  
 ax.set\_zlim((5, 55))  
  
 def lorenz\_deriv(x\_y\_z, t0, sigma=sigma, beta=beta, rho=rho):  
 """Compute the time-derivative of a Lorenz system."""  
 x, y, z = x\_y\_z  
 return [sigma \* (y - x), x \* (rho - z) - y, x \* y - beta \* z]  
  
 # Choose random starting points, uniformly distributed from -15 to 15  
 np.random.seed(1)  
 x0 = -15 + 30 \* np.random.random((N, 3))

# L1 - Resources

- Texas A&M High Performance Research Computing (HPRC)
- HPRC Microcredentials and Courses
- ACES Quick Start Guide
- ACES Portal (ACCESS)
- ACCESS Documentation
- HPRC YouTube Channel
- [help@hprc.tamu.edu](mailto:help@hprc.tamu.edu)

# NSF ACES

## Accelerating Computing for Emerging Sciences

### Our Mission:

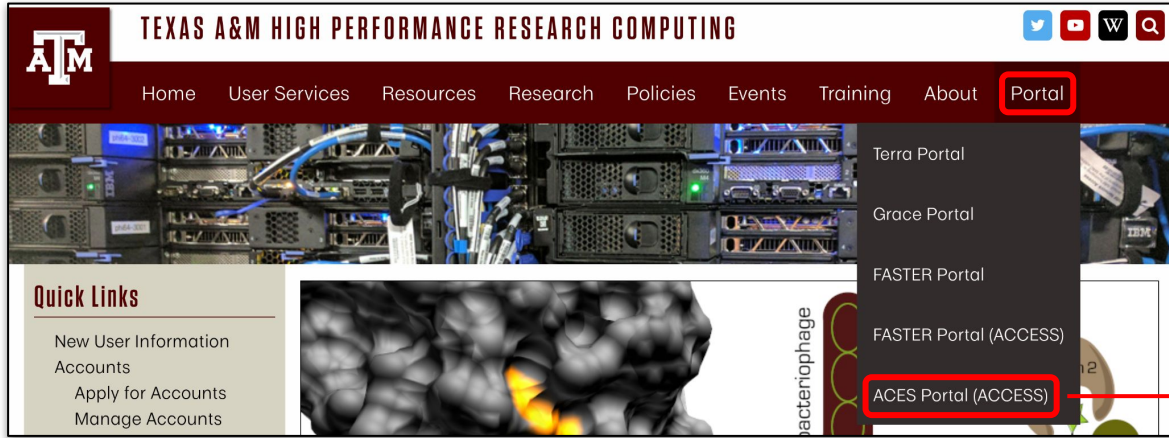
- Offer an accelerator testbed for numerical simulations and AI/ML workloads
- Provide consulting, technical guidance, and training to researchers
- Collaborate on computational and data-enabled research.



# ACES Accelerators

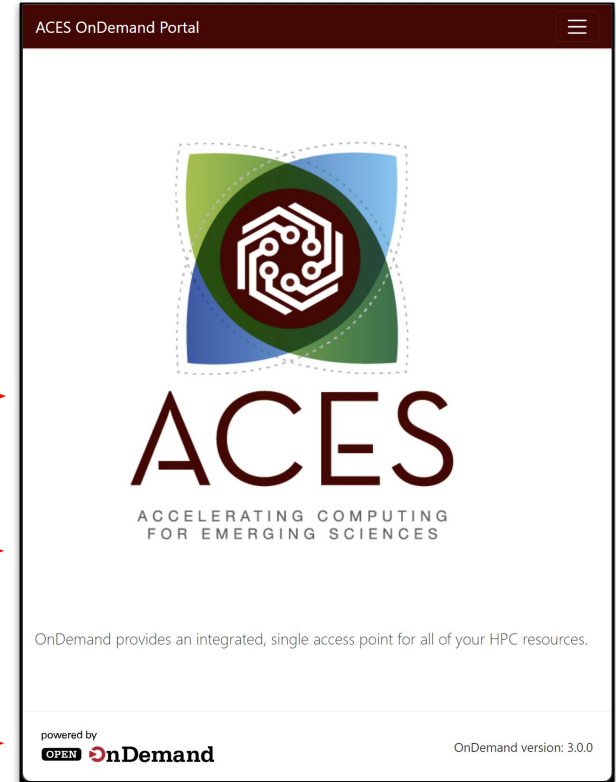
Component	Quantity	Description
Graphcore IPU	32	16 Colossus GC200 IPUs, 16 Bow IPUs. Each IPU group hosted with a CPU server as a POD16 on a 100 GbE RoCE fabric
Intel PAC D5005 FPGA	2	Accelerator with Intel Stratix 10 GX FPGA and 32 GB DDR4
BittWare IA-840F FPGA	2	Accelerator with Agilex AGF027 FPGA and 64 GB of DDR4
NextSilicon Coprocessor	2	Reconfigurable accelerator with an optimizer continuously evaluating application behavior.
NEC Vector Engine	8	Vector computing card (8 cores and HBM2 memory)
Intel Optane SSD	48	18 TB of Intel Optane SSDs addressable as memory w/ MemVerge Memory Machine.
NVIDIA H100 + A30	30 + 4	NVIDIA GPUs for HPC, DL Training, AI Inference
Intel PVC	110	Software Development Platform for PVC

# ACES Portal



ACES Portal [portal-aces.hprc.tamu.edu](http://portal-aces.hprc.tamu.edu)  
is the web-based user interface for the ACES cluster

Open OnDemand (OOD) is an advanced web-based  
graphical interface framework for HPC users



# Authentication via CILogon

Log-in using your ACCESS CI credentials.

The screenshot shows the ACCESS CI login interface. At the top, there's a 'Consent to Attribute Release' section with a dropdown arrow. Below it, a list of attributes being requested: 'Your CILogon user identifier', 'Your name', 'Your email address', and 'Your username and affiliation from your identity provider'. The main section is titled 'Select an Identity Provider' and contains a button labeled 'ACCESS CI (XSEDE)' with a question mark icon. Below this button is a 'Remember this selection' checkbox and a 'Log On' button. At the bottom, there's a link to the privacy policy.

The screenshot shows the ACCESS CI login interface. At the top, there's a 'Login to CILogon' section with input fields for 'ACCESS Username' and 'ACCESS Password'. Below these fields is a 'Don't Remember Login' checkbox and a 'Login' button. To the right of the login fields is the CILogon logo and a brief description: 'CILogon facilitates secure access to CyberInfrastructure (CI)'. Below the logo, there are links for 'Register for an ACCESS Account', 'Forgot your password?', and 'Need Help?'. At the bottom, there's a link for 'Click Here for Assistance'.

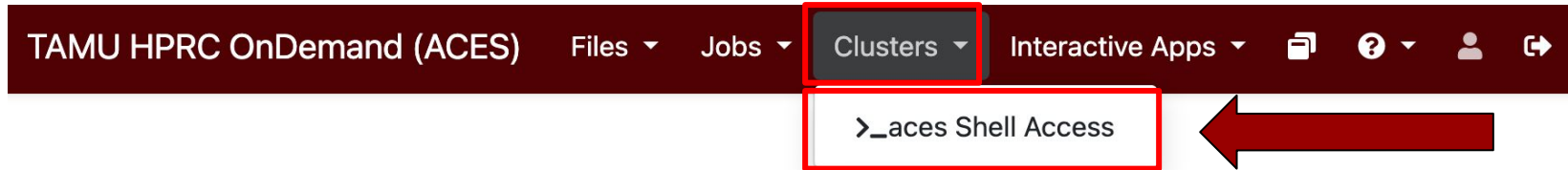
A close-up of the 'Select an Identity Provider' button, which is labeled 'ACCESS CI (XSEDE)' with a question mark icon.

Select the Identity Provider appropriate for your account.



# Get a Shell on ACES

Click on “Clusters” menu → \_aces Shell Access



# Success!

Welcome to the ACES login node.

Check which login node you are on.

```
Host: login.aces Themes: Default
| Consulting:      help@hprc.tamu.edu (preferred) or (979) 845-0219 |
| ACES Documentation: https://hprc.tamu.edu/kb/User-Guides/ACES |
| FASTER Documentation: https://hprc.tamu.edu/kb/User-Guides/FASTER |
| Grace Documentation: https://hprc.tamu.edu/kb/User-Guides/Grace |
| Terra Documentation: https://hprc.tamu.edu/kb/User-Guides/Terra |
| YouTube Channel: https://www.youtube.com/texasamhprc |
|=====|

*****
*                               *****
*                               === IMPORTANT POLICY INFORMATION ===
*                               *****
* - Unauthorized use of HPRC resources is prohibited and subject to
*   criminal prosecution.
* - Use of HPRC resources in violation of United States export control
*   laws and regulations is prohibited. Current HPRC staff members are
*   US citizens and legal residents.
* - Sharing HPRC account and password information is in violation of
*   Texas State Law. Any shared accounts will be DISABLED.
* - Authorized users must also adhere to ALL policies at:
*   https://hprc.tamu.edu/policies/
*                               *****
*                               !! WARNING: THERE ARE ONLY NIGHTLY BACKUPS OF USER HOME DIRECTORIES. !!
*
* Please restrict usage to 8 CORES across ALL login nodes.
* Users found in violation of this policy will be SUSPENDED.
*
* To see these messages again, run the motd command.
Your current disk quotas are:
Disk                               Disk Usage      Limit      File Usage      Limit
/home/u.zh108696                   4.0G           10.0G        2361            10000
/scratch/user/u.zh108696          275.4G         1.0T        352057          1000000
Type 'showquota' to view these quotas again.
[u.zh108696@aces-login1 ~]$
```

# Commands to copy the materials

- Navigate to your personal scratch directory

```
$ cd $SCRATCH
```

- Files for this course are located at

```
/scratch/training/nh_wkshop
```

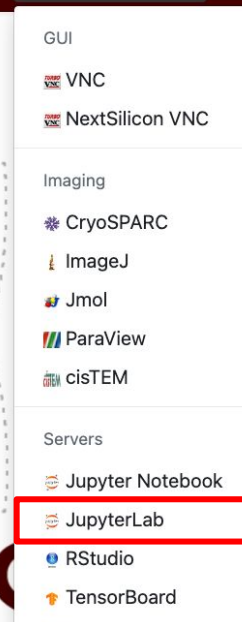
Make a copy in your personal scratch directory

```
$ cp -r /scratch/training/nh_wkshop $SCRATCH
```

- Enter this directory (your local copy)

```
$ cd nh_wkshop
```

# Go to JupyterLab Page



# JupyterLab Page

## Interactive Apps

GUI

VNC

NextSilicon VNC

Imaging

CryoSPARC

ImageJ

Jmol

ParaView

cisTEM

Servers

Jupyter Notebook

JupyterLab

## JupyterLab

This app will launch a [JupyterLab](#) server on the [ACES cluster](#).

Module

Anaconda3/2022.10

Optional Environment to be activated

/sw/hprc/sw/Anaconda3/2022.10/envs/cybertraining-er

Enter the full path and name of the environment to be activated.

Leave blank to use the [default](#) environment for the selected Module.

Your optional conda environment must have been previously built with one of the Anaconda or Python modules listed in the Module option above. See [instructions](#).

Node type

CPU only

### Other fields:

Node Type: CPU only

Number of hours: 2

Number of cores: 2

Total memory (GB): 5

**Option 1: Use a shared environment created by TAMU HPRC for this workshop**

**Path to the shared environment:**

`/sw/hprc/sw/Anaconda3/2022.10/envs/cybertraining-env`

# Connect to JupyterLab

ACES OnDemand Portal Files Jobs Clusters Interactive Apps Affinity Groups Dashboard

Session was successfully created. ✕

Home / My Interactive Sessions

Interactive Apps

GUI

VNC

NextSilicon VNC

Imaging

CryoSPARC

ImageJ

Jmol

ParaView

cisTEM

Servers

Jupyter Notebook

JupyterLab (193506) 1 node | 2 cores | Running

Host: >\_ac040 Delete

Created at: 2024-07-01 14:54:06 CDT

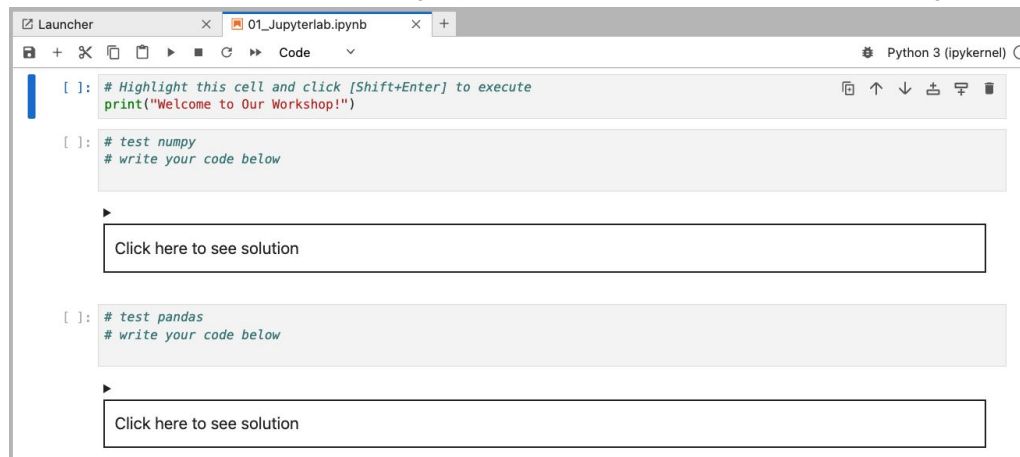
Time Remaining: 1 hour and 59 minutes

Session ID: e6d457b4-e4b8-434b-9627-43f4d2b62cc9

Connect to JupyterLab

# Review and Exercise

- Log into ACES through ACES Portal (ACCESS)
- Copy the training materials to your \$SCRATCH directory
- Launch JupyterLab app
- In the notebook named *01\_Jupyterlab.ipynb*, follow the instructions to import the required modules to make sure they have been loaded properly.



```
[ ]: # Highlight this cell and click [Shift+Enter] to execute
print("Welcome to Our Workshop!")

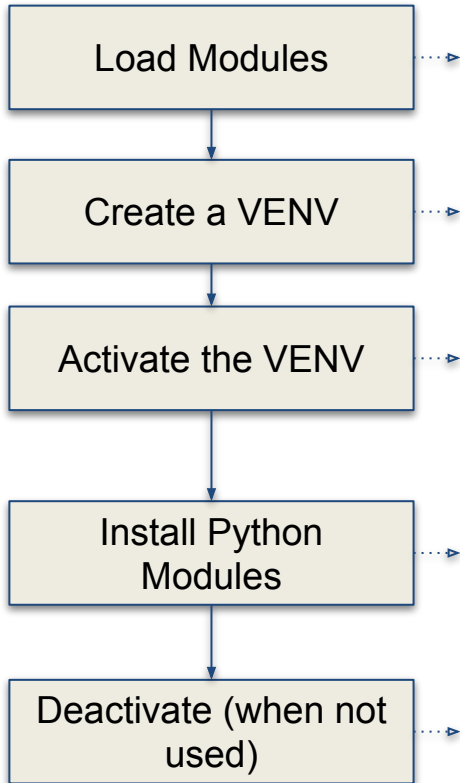
[ ]: # test numpy
# write your code below

Click here to see solution

[ ]: # test pandas
# write your code below

Click here to see solution
```

# Option 2



```
# clean up and load Anaconda  
cd $SCRATCH  
module purge  
module load Anaconda3/2022.10
```

```
# create a Python virtual environment  
conda create -n my-cybertraining-env
```

```
# activate the virtual environment  
source activate my-cybertraining-env
```

```
# install required package to be used in the portal  
conda install -c conda-forge jupyterlab  
conda install -c conda-forge numpy  
conda install -c conda-forge pandas  
conda install -c conda-forge xarray geopandas folium  
conda install -c conda-forge netcdf4  
(install other packages as well ...)
```

```
# deactivate the virtual environment  
# source deactivate
```



# JupyterLab Page

ACES OnDemand Portal Files ▾ Jobs ▾ Clusters ▾ Interactive Apps ▾ Affinity Groups ▾ Dashboard

Home / My Interactive Sessions / JupyterLab

Interactive Apps

GUI

VNC

NextSilicon VNC

Imaging

CryoSPARC

ImageJ

Jmol

ParaView

cisTEM

Servers

Jupyter Notebook

JupyterLab

## JupyterLab

This app will launch a [JupyterLab](#) server on the [ACES cluster](#).

Module

Anaconda3/2022.10

Optional Environment to be activated

my-cybertraining-env

Enter the full path and name of the environment to be activated.

Leave blank to use the [default](#) environment for the selected Module.

Your optional conda environment must have been previously built with one of the Anaconda or Python modules listed in the Module option above. See [instructions](#).

Node type

CPU only

### Other fields:

Node Type: CPU only

Number of hours: 2

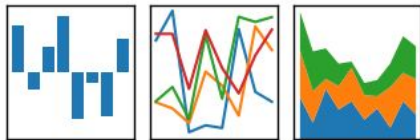
Number of cores: 2

Total memory (GB): 5

# Lab II. Data Exploration

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



matplotlib

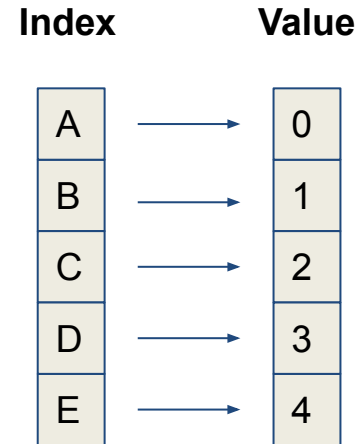
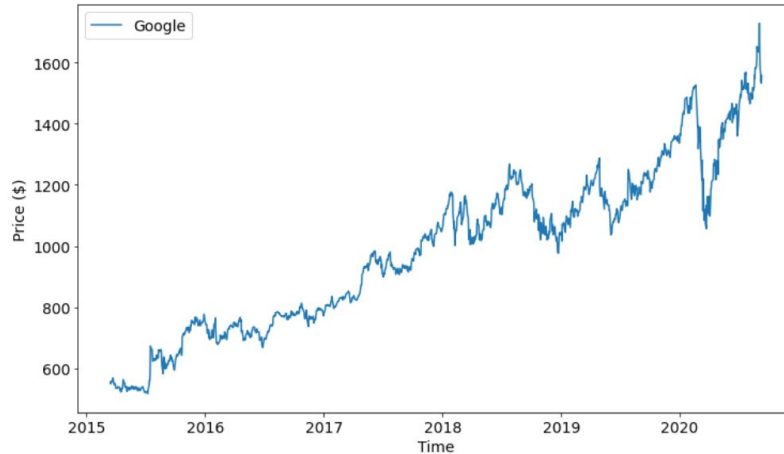
# Data Structures

**Pandas** has two data structures that are descriptive and optimized for data with different dimensions.

- **Series:** 1D labeled array
- **DataFrame:** General 2D labeled, size-mutable tabular structure with potentially heterogeneously-typed columns

# Series in pandas

- One-dimensional labeled array
- Capable of holding any data type (integers, strings, floating point numbers, etc.)
- Example: time-series stock price data



# DataFrame in pandas

- Primary Pandas data structure
- A dict-like container for Series objects
- Two-dimensional size-mutable
- Heterogeneous tabular data structure

DATE and TIME (UTC)	LAT	LON	MAG	DEPTH km
01-JUL-2024 17:10:59	13.97	52.28	4.6	10
01-JUL-2024 15:22:40	-5.51	147.18	4.5	192
01-JUL-2024 13:04:10	-32.63	-69.34	4.3	10
01-JUL-2024 11:54:40	-61.97	155.22	4.6	10
01-JUL-2024 11:37:41	-23.41	-68.10	4.5	126
01-JUL-2024 10:20:59	37.86	73.24	4.6	10
01-JUL-2024 10:16:23	-5.56	153.59	5.0	42
01-JUL-2024 05:53:52	10.79	125.37	4.8	95
01-JUL-2024 05:22:23	10.74	125.39	4.8	68
01-JUL-2024 05:12:03	14.74	147.06	5.0	10
01-JUL-2024 02:38:03	-62.68	165.56	5.1	10

(source: <https://ds.iris.edu/seismon/eventlist/index.shtml>)

Index

	C1	C2	C3	C4
A	0	x	0.1	True
B	1	y	2.4	False
C	2	z	1.9	True
D	NA	w	8.3	False
E	9	a	6.8	False

Columns

# Pandas Learning Objectives

## After this section, you will learn:

- DataFrame building
- DataFrame operations
  - Relabeling
  - Data grouping
- Data handling
  - Handle missing data
  - Handle duplicate data
  - Merge DataFrames



**JupyterLab Exercises**

# Lab III. Machine/Deep Learning

## ***Deep Learning***

by Ian Goodfellow, Yoshua Bengio, and Aaron Courville

<http://www.deeplearningbook.org/>

## ***Animation of Neutron Networks***

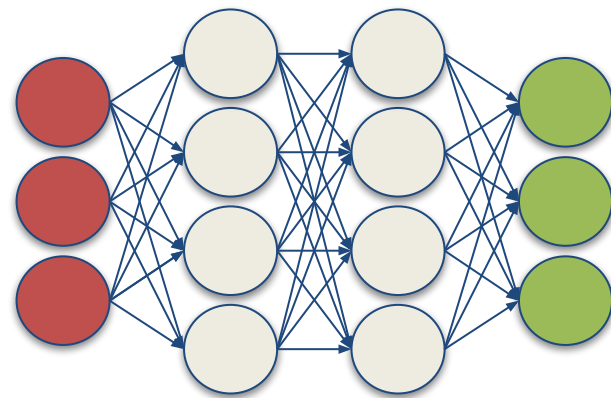
by Grant Sanderson

<https://www.3blue1brown.com/>

## ***Visualization of CNN***

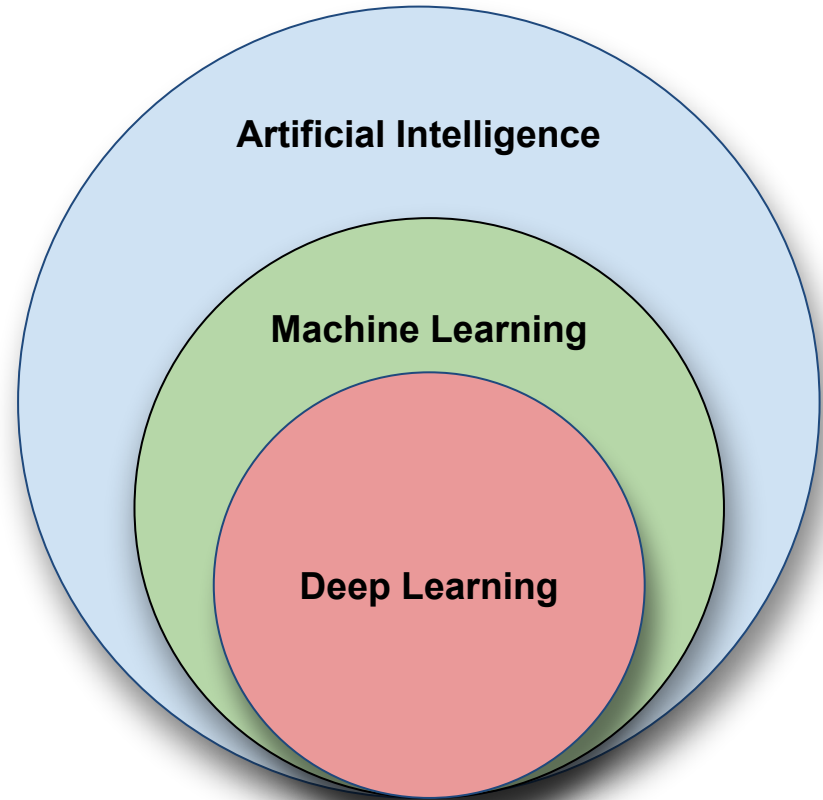
by Adam Harley

[https://adamharley.com/nn\\_vis/cnn/3d.html](https://adamharley.com/nn_vis/cnn/3d.html)



# Relationship of AI, ML, and DL

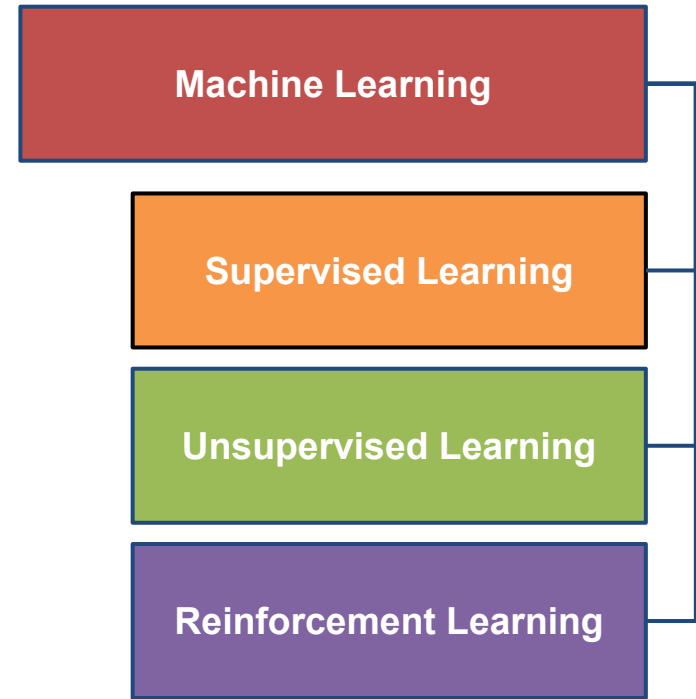
- **Artificial Intelligence (AI)** is anything about man-made intelligence exhibited by machines.
- **Machine Learning (ML)** is an approach to achieve **AI**.
- **Deep Learning (DL)** is one technique to implement **ML**.





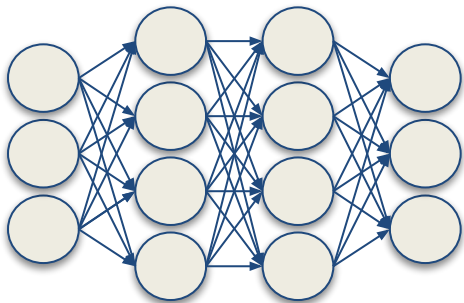
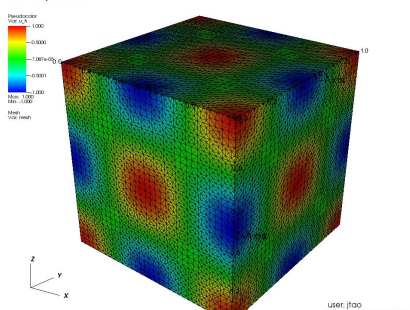
# Types of ML Algorithms

- **Supervised Learning**
  - trained with labeled data; including regression and classification problems
- **Unsupervised Learning**
  - trained with unlabeled data; clustering and association rule learning problems.
- **Reinforcement Learning**
  - no training data; stochastic Markov decision process; robotics and business strategy planning.

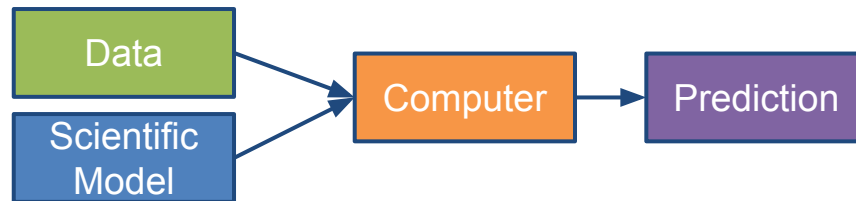


# Machine Learning

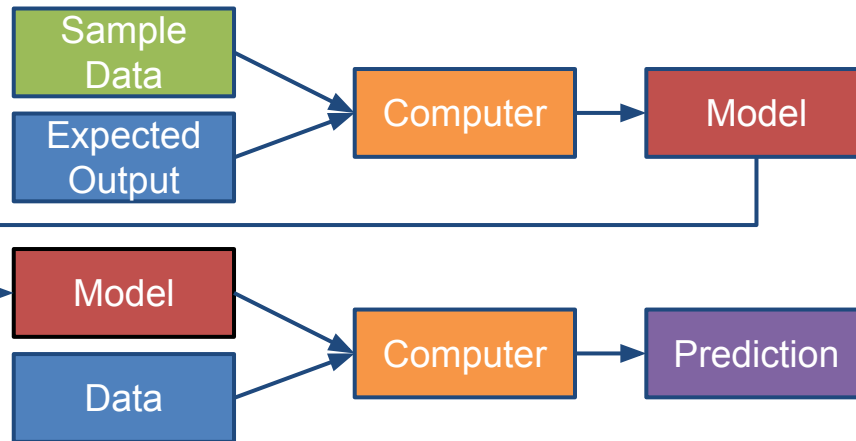
DB: simplest.vtk



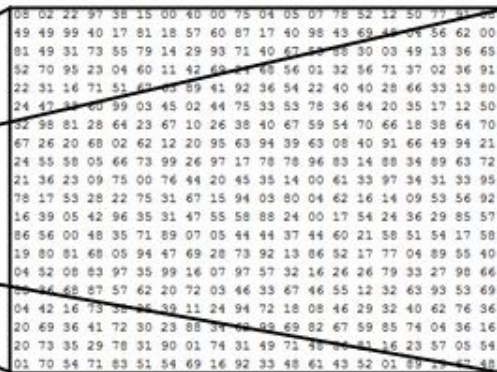
## Traditional Modeling



## Machine Learning (Supervised Learning)



# Inputs and Outputs



What the computer sees

image classification

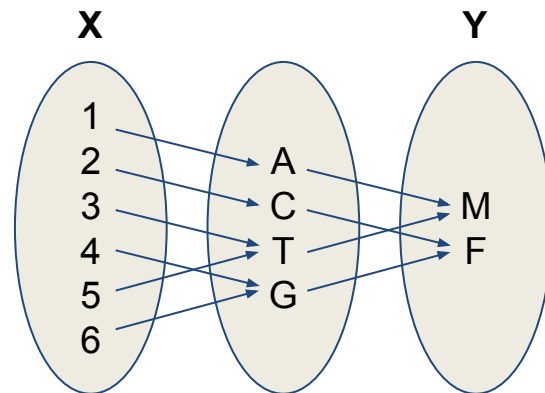
82% cat  
15% dog  
2% hat  
1% mug

Image from the [Stanford CS231 Course](#)

256 X 256  
Matrix

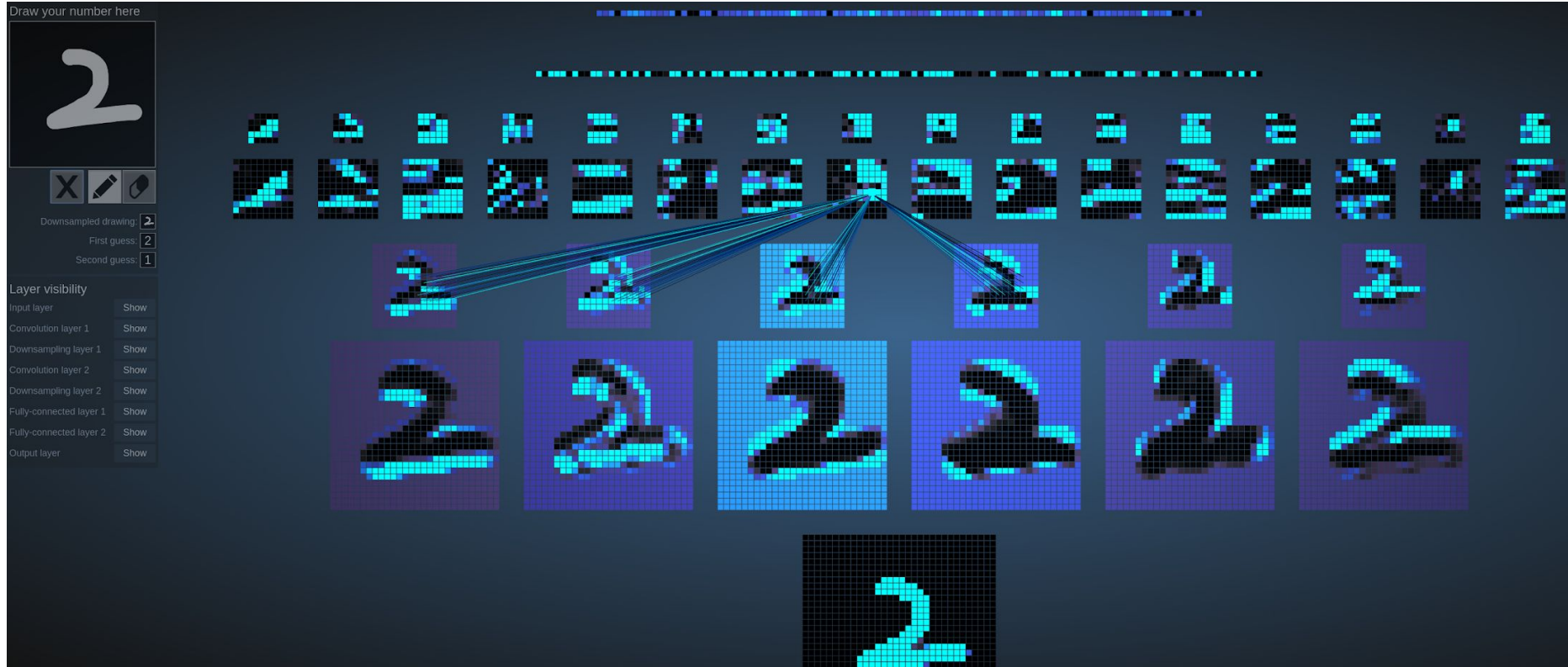
DL model

4-Element Vector



With deep learning, we are searching for a **surjective** (or **onto**) function  $f$  from a set  $X$  to a set  $Y$ .

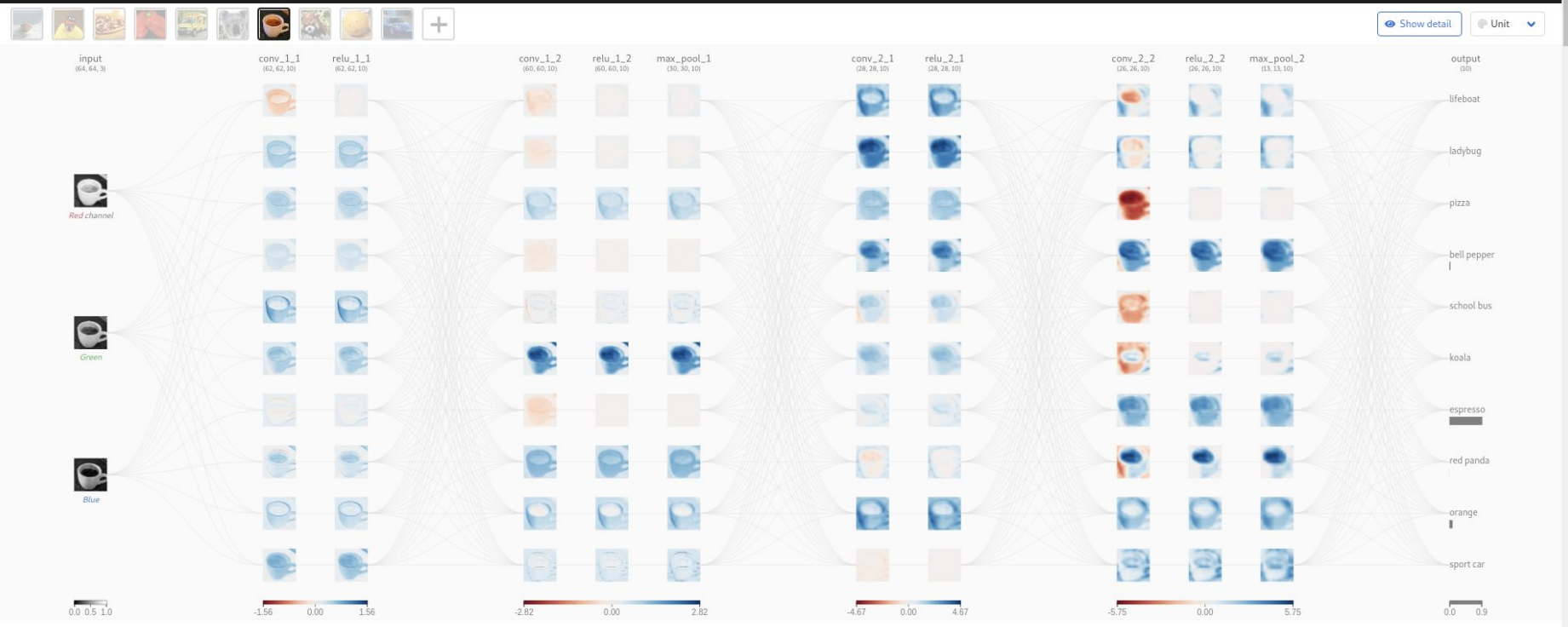
# MNIST - CNN Visualization



(Image Credit: [https://adamharley.com/nn\\_vis/cnn/3d.html](https://adamharley.com/nn_vis/cnn/3d.html))

# CNN Explainer

**CNN EXPLAINER** Learn Convolutional Neural Network (CNN) in your browser!



(Image Credit: <https://poloclub.github.io/cnn-explainer/>)



**JupyterLab Exercises**



# High Performance Research Computing

**DIVISION OF RESEARCH**

<https://hprc.tamu.edu>

HPRC Helpdesk:

help@hprc.tamu.edu

Phone: 979-845-0219

Help us help you. Please include details in your request for support, such as, Cluster (ACES, FASTER, Grace, Terra, ViDaL), NetID (UserID), Job information (Job id(s), Location of your jobfile, input/output files, Application, Module(s) loaded, Error messages, etc), and Steps you have taken, so we can reproduce the problem.