

Tutorial: Python for Data Science on **ACES**

Richard Lawrence
2025/04/11



High Performance
Research Computing
DIVISION OF RESEARCH



Outline

- Overview of Python for Data Science
- Getting Started
- Arrays and Series
- Plots

Overview of Python for Data Science

Arrays

Numpy Arrays support common operations, such as arithmetic, on an element-by-element (or “vectorized”) basis.

Example:

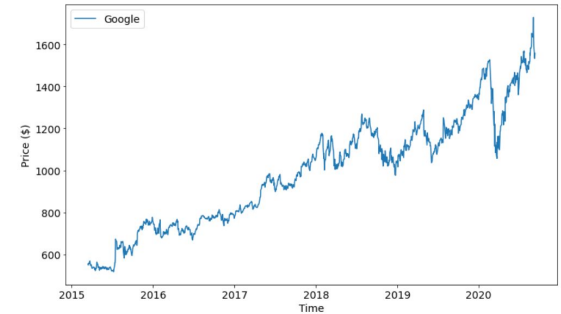
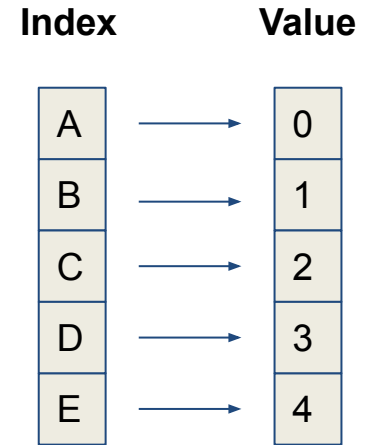
```
array C = array A + array B
```

This adds the elements of A and B pair-wise (Instead of concatenating the elements as would happen with lists).

Pandas Series and DataFrames further expand on this.

Series

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



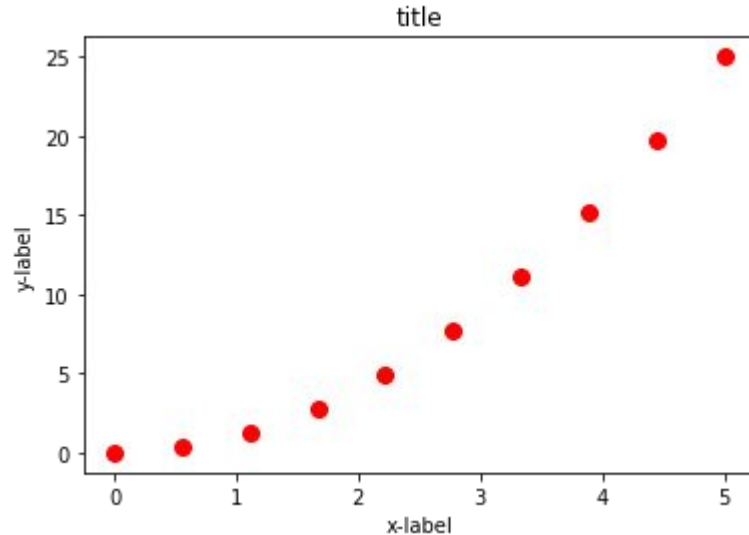
Anatomy of a Plot

Marker

- style
- size
- color

Figure

- title
- xlabel
- ylabel

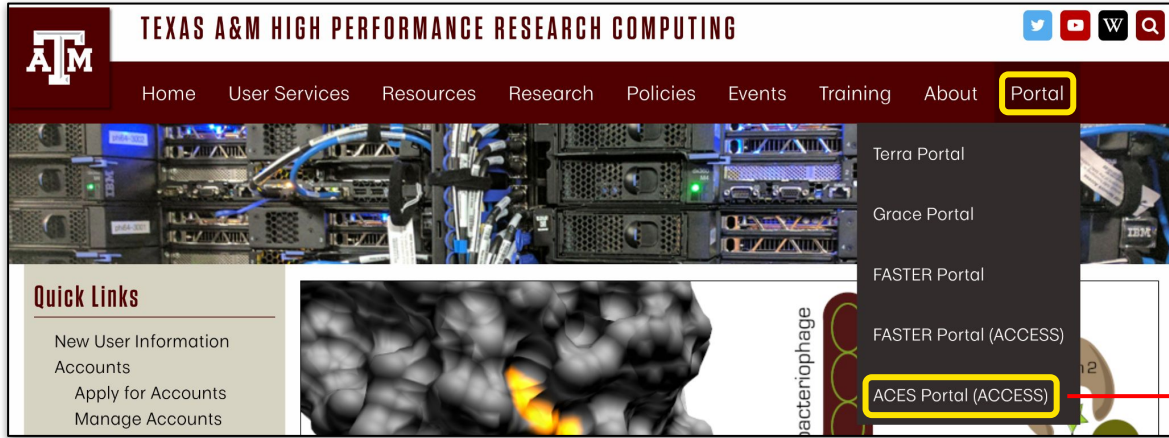


Learning Resources

- Slides on the course web page
https://hprc.tamu.edu/training/aces_python4data.html
- HPRC's Knowledge Base
<https://hprc.tamu.edu/kb/Software/Python/>
- HPRC on YouTube
<https://www.youtube.com/c/TexasAMHPRC>
- ACCESS Links
<https://support.access-ci.org/ci-links>

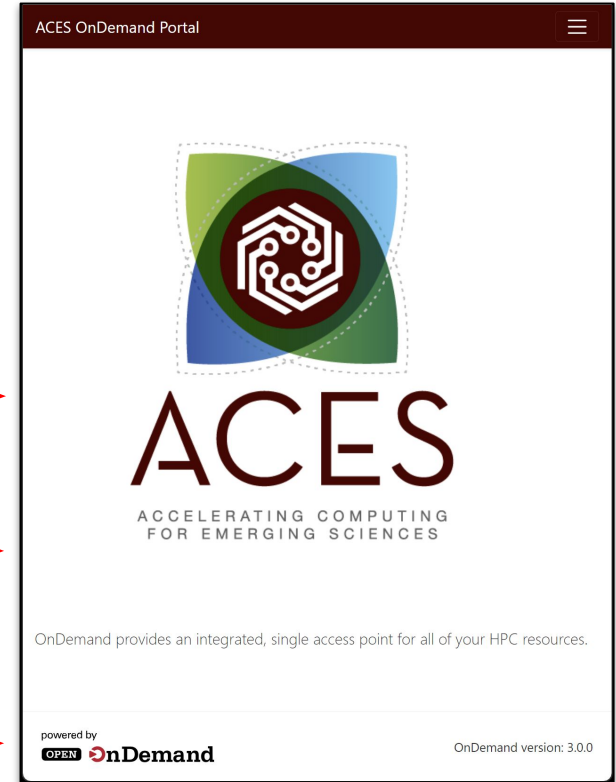
Getting Started

ACES Portal

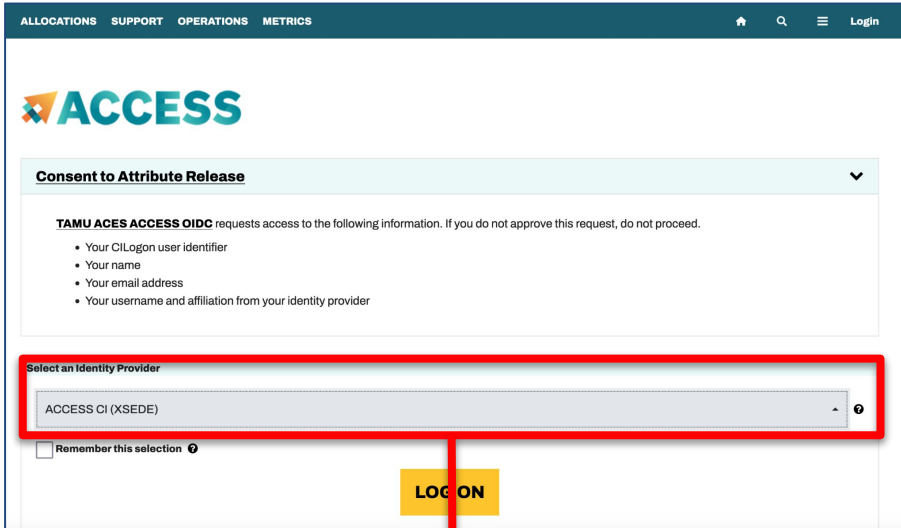


ACES Portal portal-aces.hprc.tamu.edu
is the web-based user interface for the ACES cluster

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

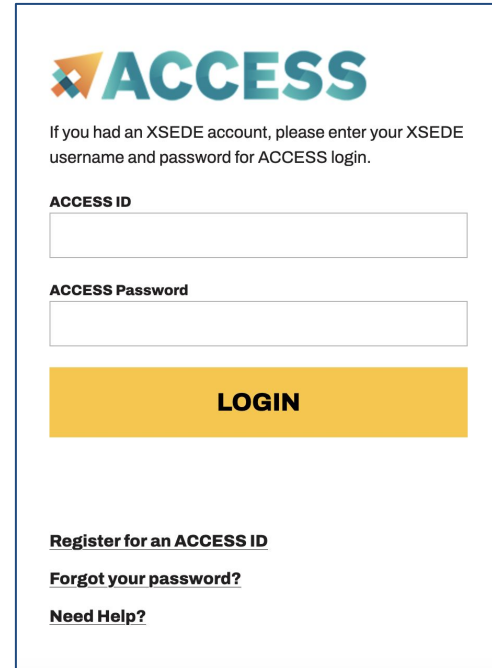


Accessing ACES via the Portal (ACCESS)



The screenshot shows the ACCESS portal interface. At the top is a navigation bar with links: ALLOCATIONS, SUPPORT, OPERATIONS, METRICS, and a Login link. Below the navigation bar is the ACCESS logo. A section titled "Consent to Attribute Release" is expanded, showing a message from TAMU ACES ACCESS OIDC requesting access to user information. Below this is a "Select an Identity Provider" dropdown menu, which is highlighted with a red rectangle. The dropdown shows "ACCESS CI (XSEDE)" as the selected option. Below the dropdown is a checkbox labeled "Remember this selection". A yellow "LOG ON" button is positioned below the dropdown menu.

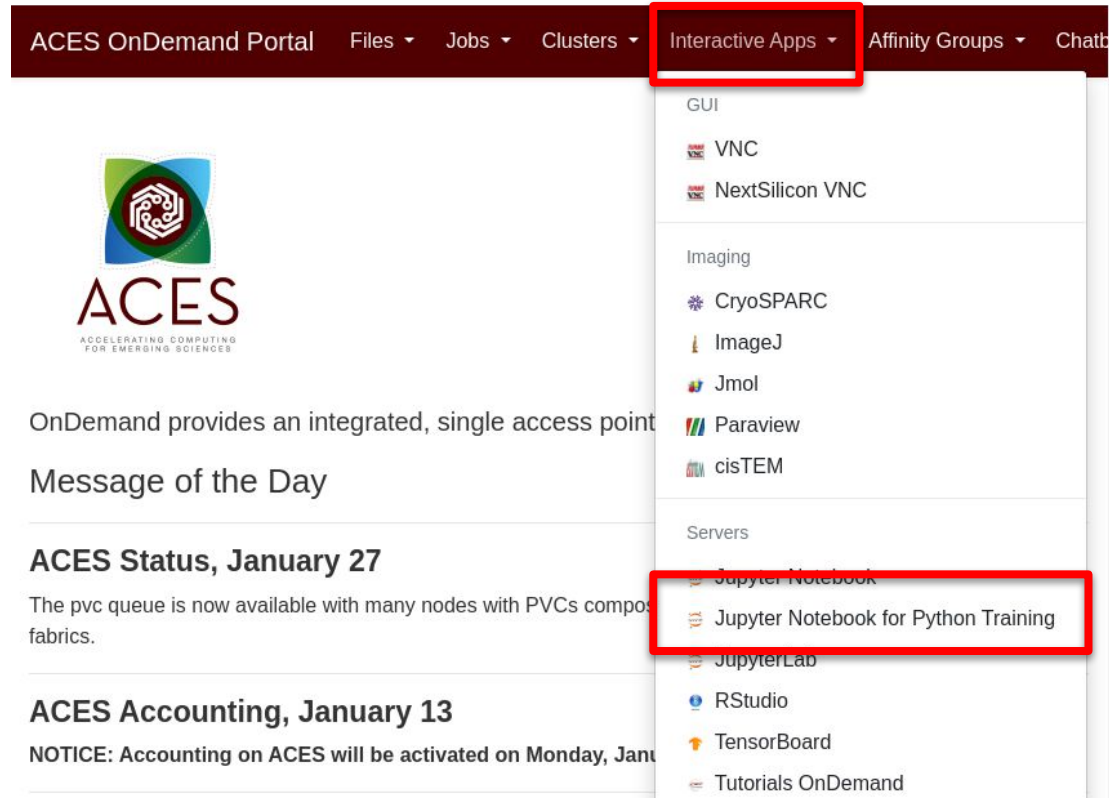
Select the Identity Provider appropriate for your account.



The screenshot shows the ACCESS portal login page. At the top is the ACCESS logo. Below the logo is a message: "If you had an XSEDE account, please enter your XSEDE username and password for ACCESS login." Below this message are two input fields: "ACCESS ID" and "ACCESS Password". Below the input fields is a large yellow "LOGIN" button. At the bottom of the page are three links: "Register for an ACCESS ID", "Forgot your password?", and "Need Help?".

Log-in using your ACCESS or institutional credentials.

Jupyter Notebook for Python Training



The screenshot displays the ACES OnDemand Portal interface. The top navigation bar includes links for 'Files', 'Jobs', 'Clusters', 'Interactive Apps', 'Affinity Groups', and 'Chat'. The 'Interactive Apps' menu is open, showing a list of applications categorized under 'GUI', 'Imaging', and 'Servers'. The 'Jupyter Notebook for Python Training' option is highlighted with a red box. The main content area on the left features the ACES logo, a message about the OnDemand portal, and status updates for January 27 and January 13.

ACES OnDemand Portal Files ▾ Jobs ▾ Clusters ▾ **Interactive Apps ▾** Affinity Groups ▾ Chat ▾

GUI

- VNC
- NextSilicon VNC

Imaging

- CryoSPARC
- ImageJ
- Jmol
- Paraview
- cisTEM

Servers

- Jupyter Notebook
- Jupyter Notebook for Python Training**
- JupyterLab
- RStudio
- TensorBoard
- Tutorials OnDemand

ACES
ACCELERATING COMPUTING
FOR EMERGING SCIENCES

OnDemand provides an integrated, single access point

Message of the Day

ACES Status, January 27

The pvc queue is now available with many nodes with PVCs composed of fabrics.

ACES Accounting, January 13

NOTICE: Accounting on ACES will be activated on Monday, Jan 13

Jupyter Notebook for Python Training

Jupyter Notebook for Python Training

This app will launch a Jupyter Notebook server on the ACES cluster.

Email

This field is optional.

Launch

* The Jupyter Notebook for Python Training session data for this session can be accessed under the [data root directory](#).

Jupyter Notebook for Python Training (636884)

1 node | 1 core | Starting

Create

Time

Session

Your s

take a

Jupyter Notebook for Python Training (636884)

1 node | 1 core | Running

Host: >_ ac058

Delete

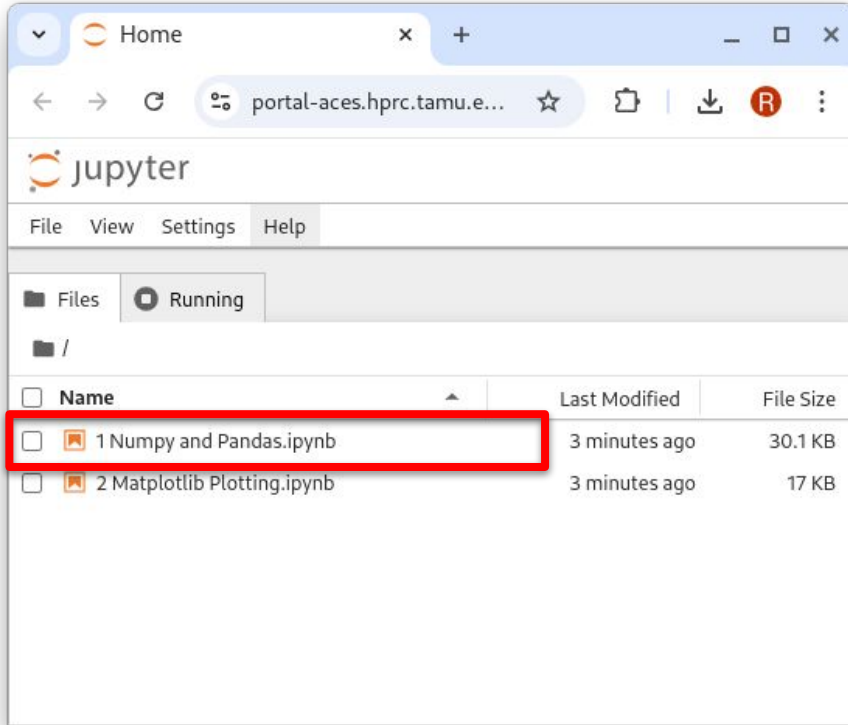
Created at: 2025-01-28 09:19:34 CST

Time Remaining: 7 hours and 9 minutes

Session ID: 428342db-6782-4d59-8bf8-2537d8ee857d

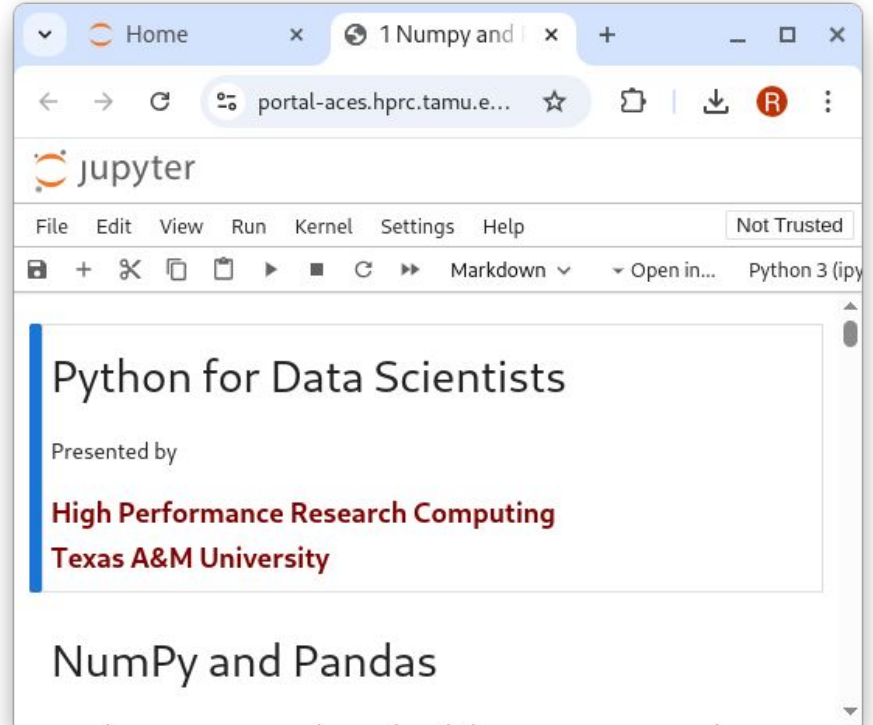
Connect to Jupyter

Jupyter Notebook for Python Training



A screenshot of the JupyterLab interface showing the file browser. The browser displays a list of files in a table with columns for Name, Last Modified, and File Size. The file "1 Numpy and Pandas.ipynb" is highlighted with a red rectangle.

<input type="checkbox"/>	Name	Last Modified	File Size
<input type="checkbox"/>	1 Numpy and Pandas.ipynb	3 minutes ago	30.1 KB
<input type="checkbox"/>	2 Matplotlib Plotting.ipynb	3 minutes ago	17 KB



A screenshot of a Jupyter Notebook titled "Python for Data Scientists". The notebook content includes the title, the presenter "High Performance Research Computing Texas A&M University", and the topic "NumPy and Pandas".

Python for Data Scientists

Presented by

High Performance Research Computing
Texas A&M University

NumPy and Pandas



**HIGH PERFORMANCE
RESEARCH COMPUTING**
TEXAS A&M UNIVERSITY

<https://hprc.tamu.edu>

HPRC Helpdesk:

help@hprc.tamu.edu

Phone: 979-845-0219

Take our short course survey!



HPRC Survey

https://u.tamu.edu/hprc_shortcourse_survey

Help us help you. Please include details in your request for support, such as, Cluster (ACES, FASTER, Grace, Launch), NetID (UserID), Job information (JobID(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.

