

# ARM Data Workbench & JupyterHub

**Kyle Dumas**  
Oak Ridge National Laboratory

**Maxwell Grover**  
Argonne National Laboratory



Kyle Dumas  
Data Tools Lead



Will Provenza  
System Admin



Wade Darnell  
User Tools Lead



Michael Giansiracusa  
DevOps Lead



Sujata Goswami  
DevOps / HPC



Max Grover  
Open Source Guru



## A Quick Note

If you run into issues logging in to JupyterHub, there are representatives from the ARM Data Booth ready to assist.

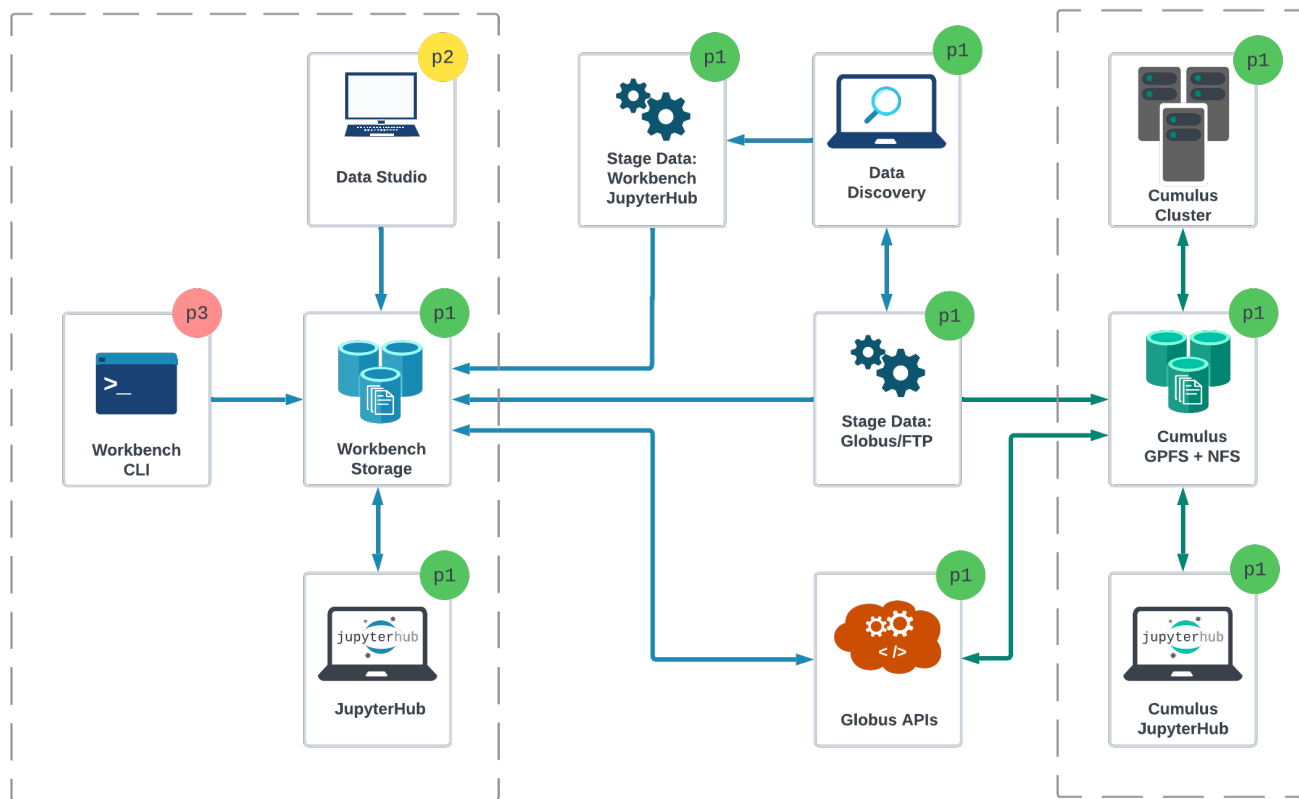
# What is the ARM Data Workbench

The ARM Data Workbench (ADW) is a revolutionary ecosystem for interacting with ARM data. It is intended to utilize the ARM Data Center's (ADC) full capabilities to improve users' experience in working with ARM data. A few of these capabilities include:

- ▶ Seamless access to data and computing resources
- ▶ Direct access to ARM and community-developed software packages
- ▶ Jupyter notebook ecosystem for data access, analysis and sub-setting
- ▶ Easy generation of standard and custom plots
- ▶ Integration of external data sets

# ARM Data Workbench Timeline

## ARM Data Workbench



### Phase 1: (COMPLETE)

JupyterLab integration with Discovery and ordering

### Phase 2:

Initial UI to support querying, filtering, and data management functionality

### Phase 3:

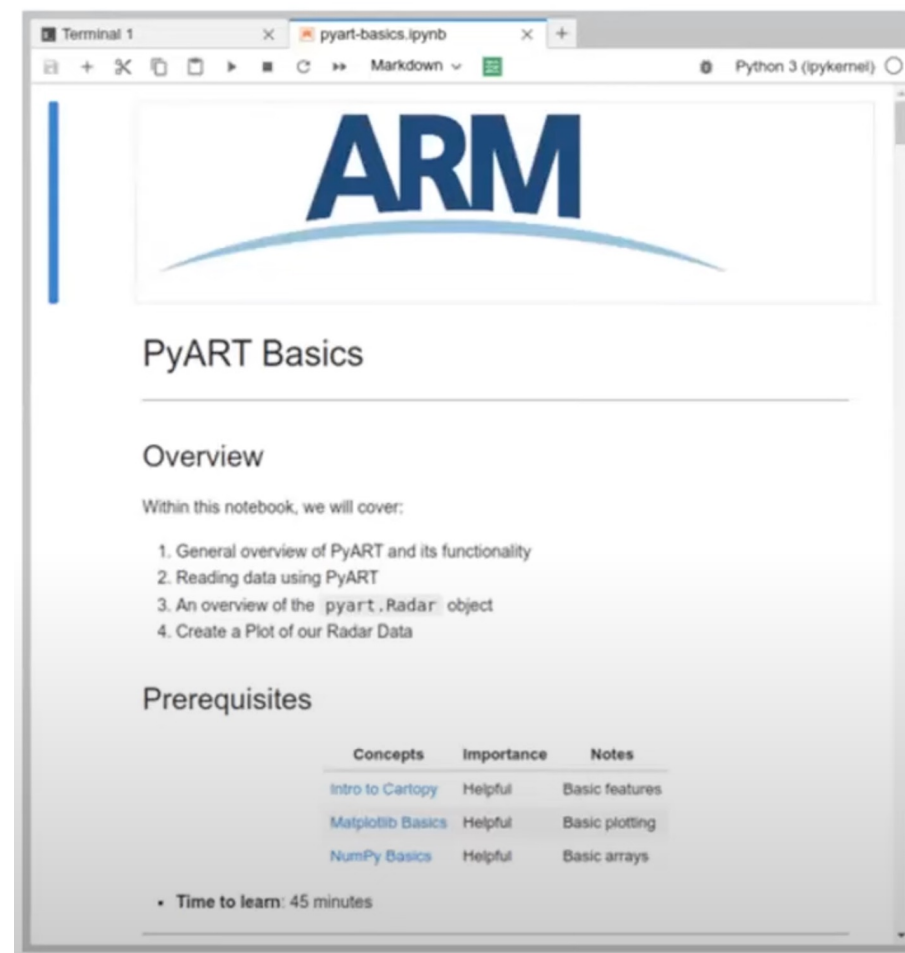
Expanding support for APIs and CLI capabilities

p1 Phase 1    
 p2 Phase 2    
 p3 Phase 3

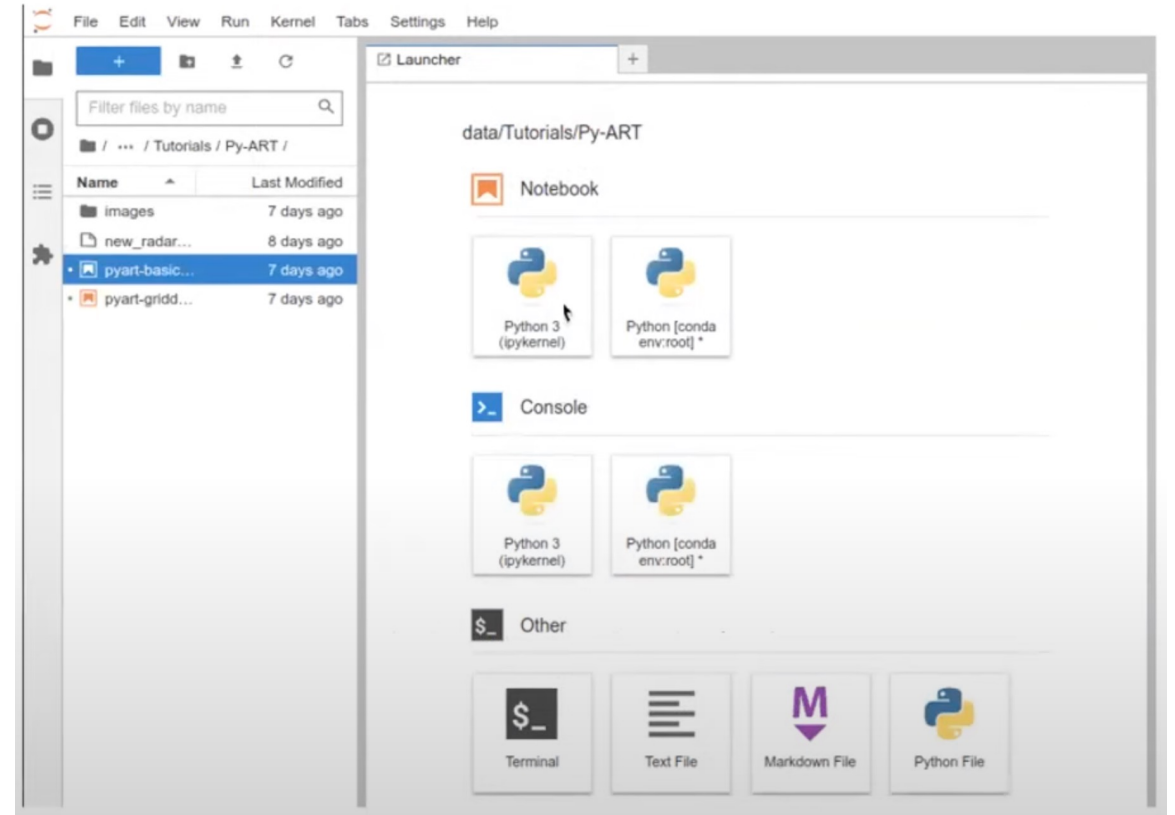
Note that Phase 1 indicates the delivery of the minimum viable product (MVP). Additional features beyond the MVP will continue to be in development beyond the initial Phase 1 delivery.

# Jupyter Notebook

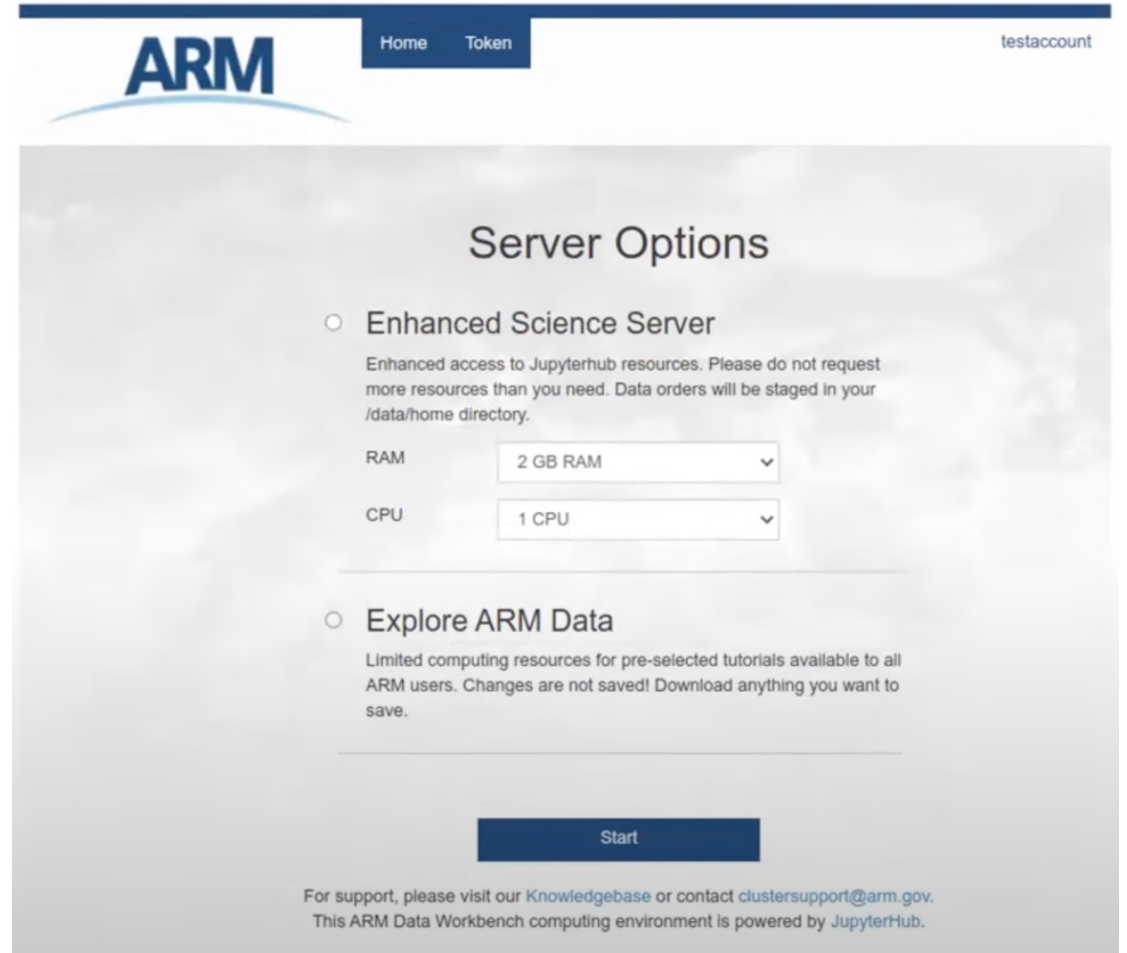
- ▶ A development environment for creating and sharing computational documents.
- ▶ Most commonly known for Python, but can support other languages as well.
- ▶ Interactive code.
- ▶ Easy to see and communicate how a program unfolds since the outcome of each cell/block is cached.
- ▶ The “classic” notebook interface for editing documents.
- ▶ Notebooks are the individual files that can be created, edited, and shared. (.ipynb files)



- ▶ A web-based interactive development environment for notebooks.
- ▶ Includes file system navigation, the ability to edit notebooks, open a console, and terminal options.
- ▶ Note that JupyterLab and Jupyter Notebooks can be installed on your local machine.



- ▶ A application for making JupyterLab and notebooks available in a multi-tenant environment.
- ▶ Each user spawns their own private JupyterLab web server.
- ▶ Benefits include:
  - Don't have to setup environment.
  - Give users consistent access to same dependencies and packages.
  - For ARM users, mount personal and shared project storage from NFS.
  - Integration with ARM Data Discovery.



The screenshot shows the ARM JupyterHub interface. At the top, there is a navigation bar with the ARM logo, 'Home', 'Token', and 'testaccount'. The main content area is titled 'Server Options' and features two radio button options:

- Enhanced Science Server**  
Enhanced access to Jupyterhub resources. Please do not request more resources than you need. Data orders will be staged in your /data/home directory.  
RAM: 2 GB RAM (dropdown)  
CPU: 1 CPU (dropdown)
- Explore ARM Data**  
Limited computing resources for pre-selected tutorials available to all ARM users. Changes are not saved! Download anything you want to save.

A 'Start' button is located at the bottom of the options section. Below the button, there is a footer with support information: 'For support, please visit our Knowledgebase or contact clustersupport@arm.gov. This ARM Data Workbench computing environment is powered by JupyterHub.'

## Service Levels

### **Explore ARM Data (Everyone)**

Default access with limited resources without persistent file storage.

### **Enhanced Science Server (Requestable)**

Scalable resources with persistent private & shared project spaces.

### **Research System Server (ARM Infrastructure Users)**

Same as the enhanced with additional mounts to internal resources (datastream, archive).

### **Workshop & Tutorial Access (NEW)**

A custom implementation of JupyterHub tailored to the needs of workshops. Instructors can get access to participants JupyterLab instances.



# How to Access These Resources

## ▶ Directly

<https://jupyterhub.arm.gov>

\*Workshops have a dedicated link

## ▶ Data Discovery

<https://adc.arm.gov/discovery>

- Homepage
- When Ordering
- Your Account

## ▶ ARM.gov

<https://arm.gov>

Go to Capabilities > Computing Resources >  
JupyterHub & ARM Data Workbench

The screenshot shows the 'Data Selection Summary' interface for the 'SONDEWNP' dataset. The interface is divided into several sections:

- Dataset Information:**
  - SONDEWNP:** Balloon-borne sounding system (BBSS); Vaisala-processed winds, press., temp, &RH
  - Data Level:** b1
  - Site:** ACAPEX (ARM Cloud Aerosol Precip Experiment); Mobile Facility (ACX)
  - Facility:** Off the Coast of California - NOAA Ship Ronald H. Brown; AMF2 (M1)
  - Category:** Atmospheric Profiling
  - Data Type:** Routine Data
  - Source Instrument/Data:** Balloon-Borne Sounding System
  - Date Range:** 2015-01-12 to 2015-02-10
- Ordering Options:**
  - Order all Variables
  - Extract Requested Variables

Note: all variables will be delivered for this datastream. Extraction options only apply when "Extract Requested Variables" is selected.
- Citation Format:** Select (dropdown menu)
- File Count:** 175 files // 41.1 MB
- Concatenate files by variable:** No (dropdown menu)
- File format(s):**
  - NetCDF
  - ASCII-CSV
- Filter data flagged by Data Quality Reports:**
  - Incorrect
  - Suspect
- Data Delivery Options:**
  - All orders are provided via FTP
  - Globus
  - THREDDS
  - Dropbox
  - NEW Stage data to Workbench JupyterHub** (Alpha)

Buttons at the bottom: Cancel, Submit Data Request

# Demo Time



# Thank You!



If you need any assistance, please stop by the ARM Data Booth and we will be happy to help.

You can also reach us by email:

[clustersupport@arm.gov](mailto:clustersupport@arm.gov)

[adc@arm.gov](mailto:adc@arm.gov)