Jupyter Docker Installation Guide

Introduction

- We will use JupyterLab as a web-based interactive development environment designed for efficient data exploration, analysis, and visualization.
- We will install JupyterLab via Docker.
    - Docker is a platform that allows you to run applications in isolated *containers (*a standalone environment that includes everything needed to run an application*)*. By installing JupyterLab via Docker, we can easily get started without worrying about software installation and configuration. This provides a streamlined workflow for data scientists and researchers.
- Learn more about:
    - JupyterLab: [link]
    - Docker: [link]

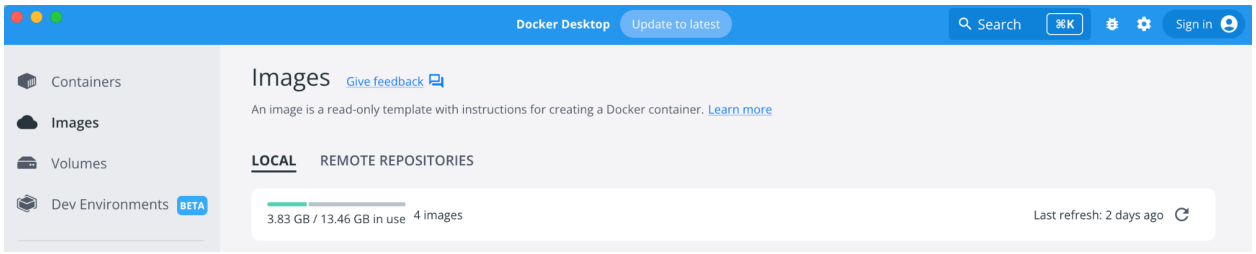
Specification

- 10GB free storage space.
- 4GB RAM.

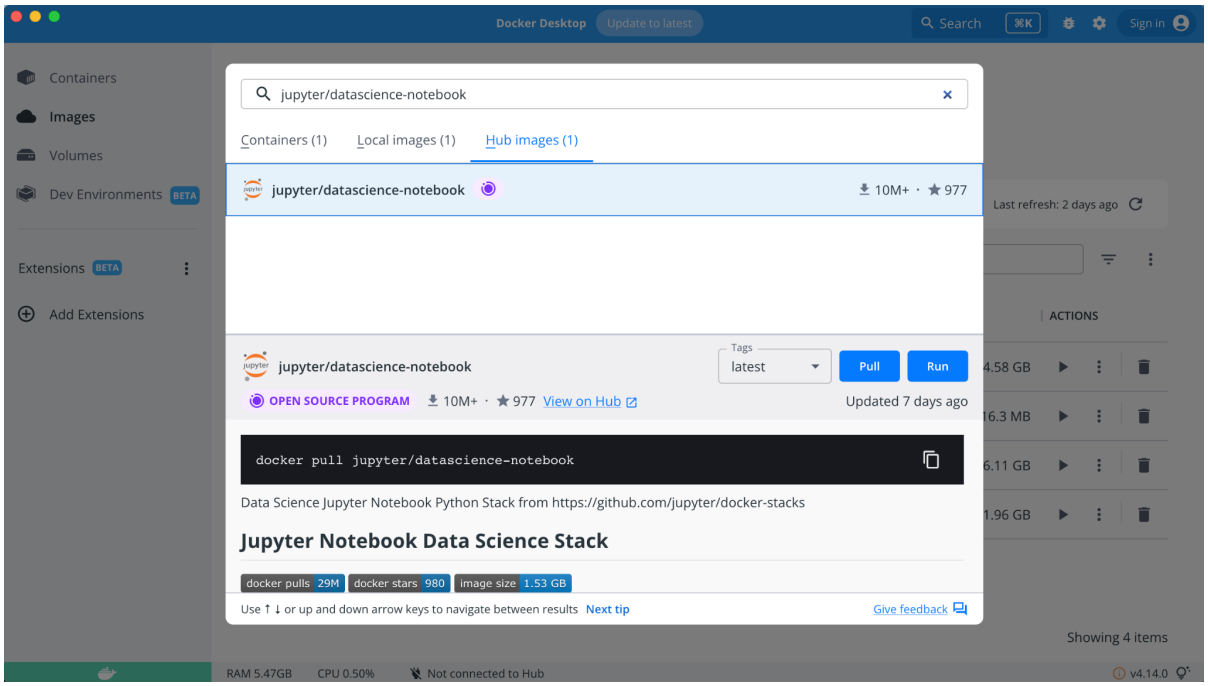
Installation Guide

1. Install Docker Desktop https://www.docker.com/products/docker-desktop/

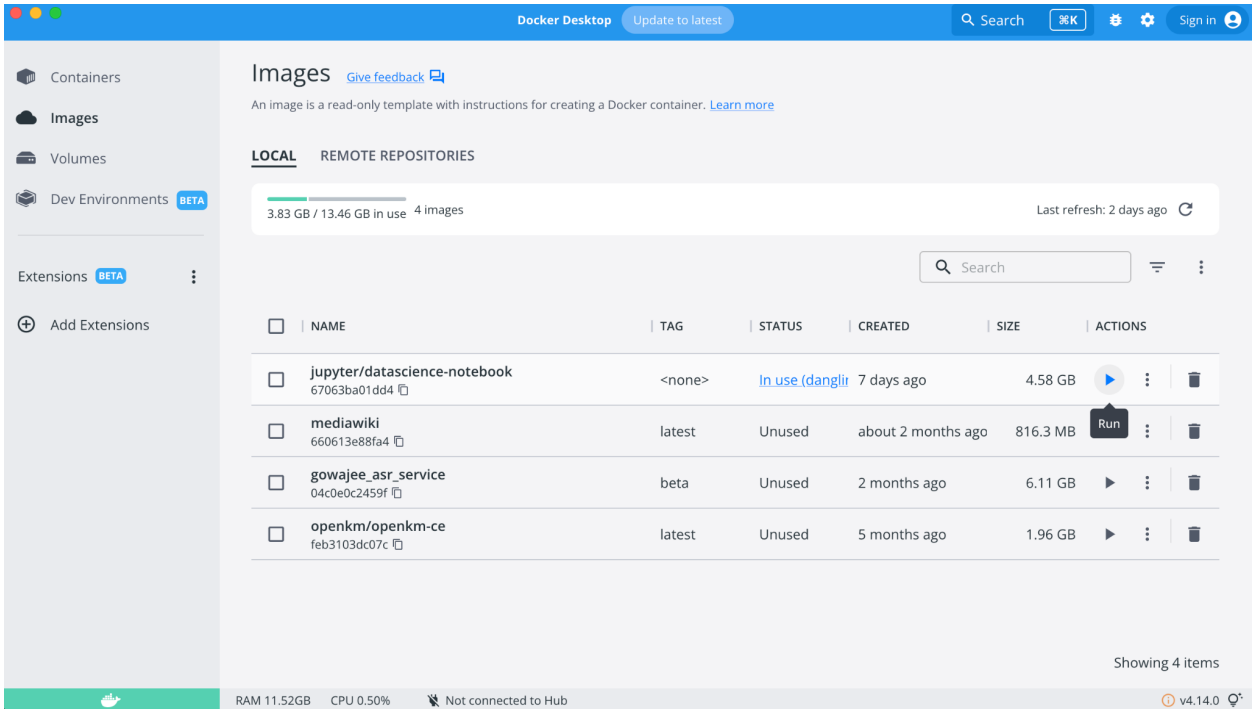
2. Open Docker Desktop Application and click on the search bar at the top right.



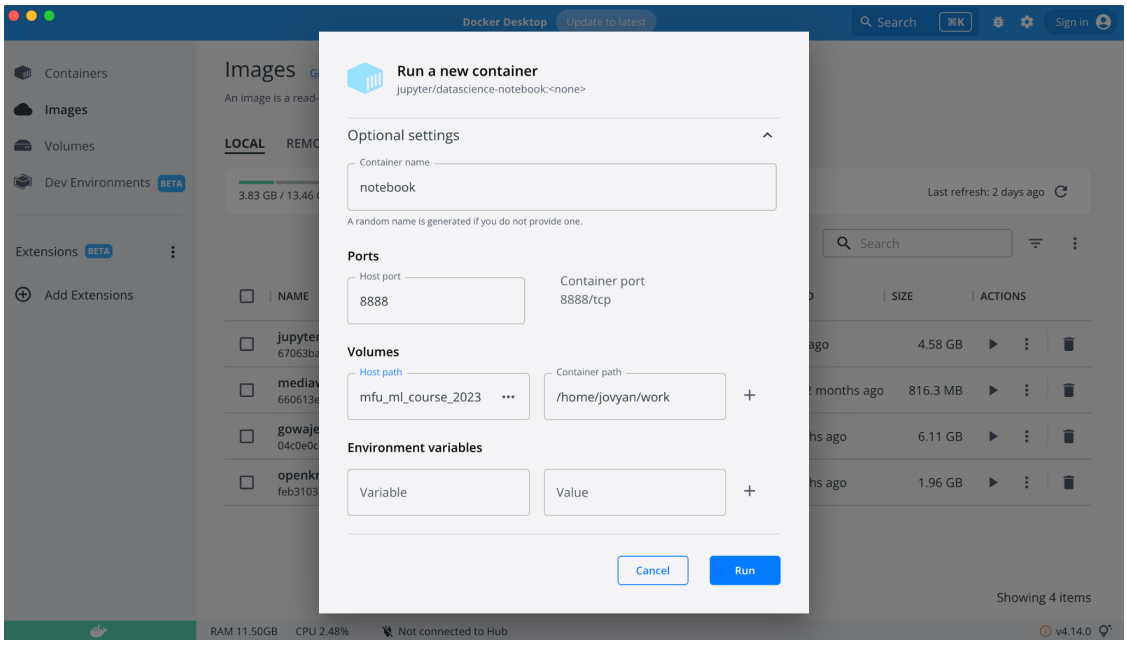
3. Type "**jupyter/datascience-notebook**" and click the **Hub Images** tab. Then, click the **Pull** button to download the Docker image onto your computer.



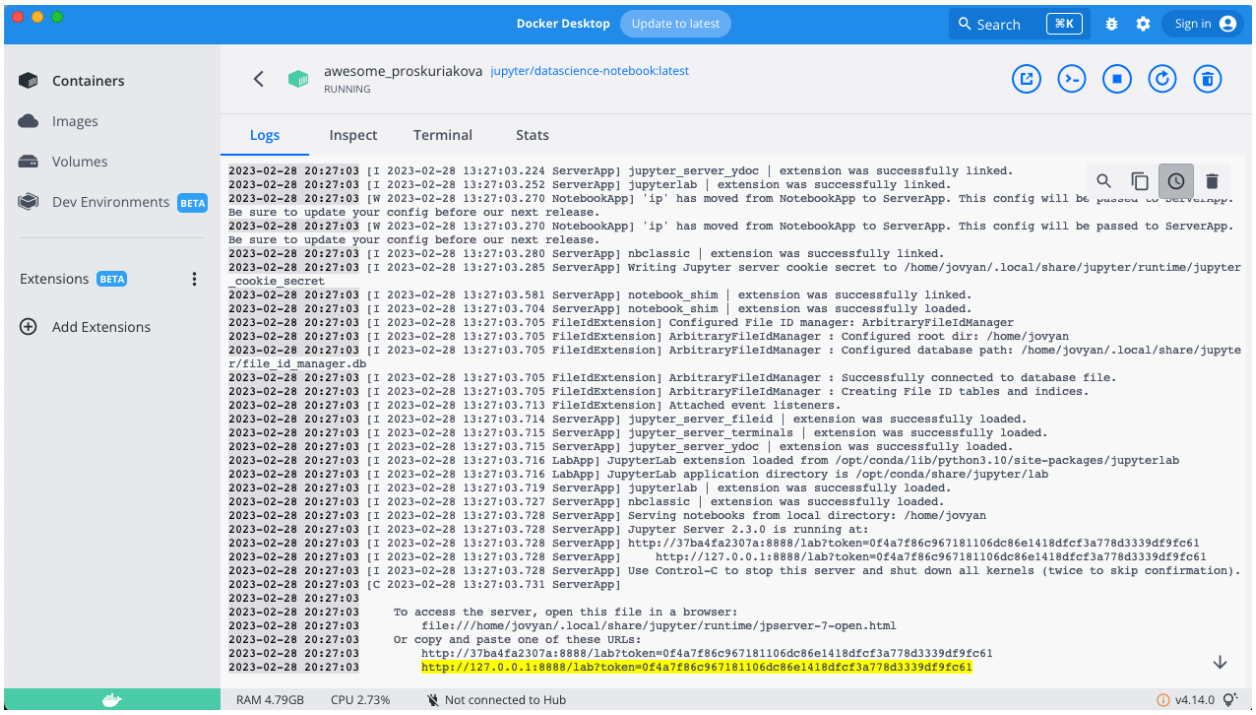
4. Click the **Image** tab on the left menu and click the **Run** button on the jupyter/datascience-notebook image.



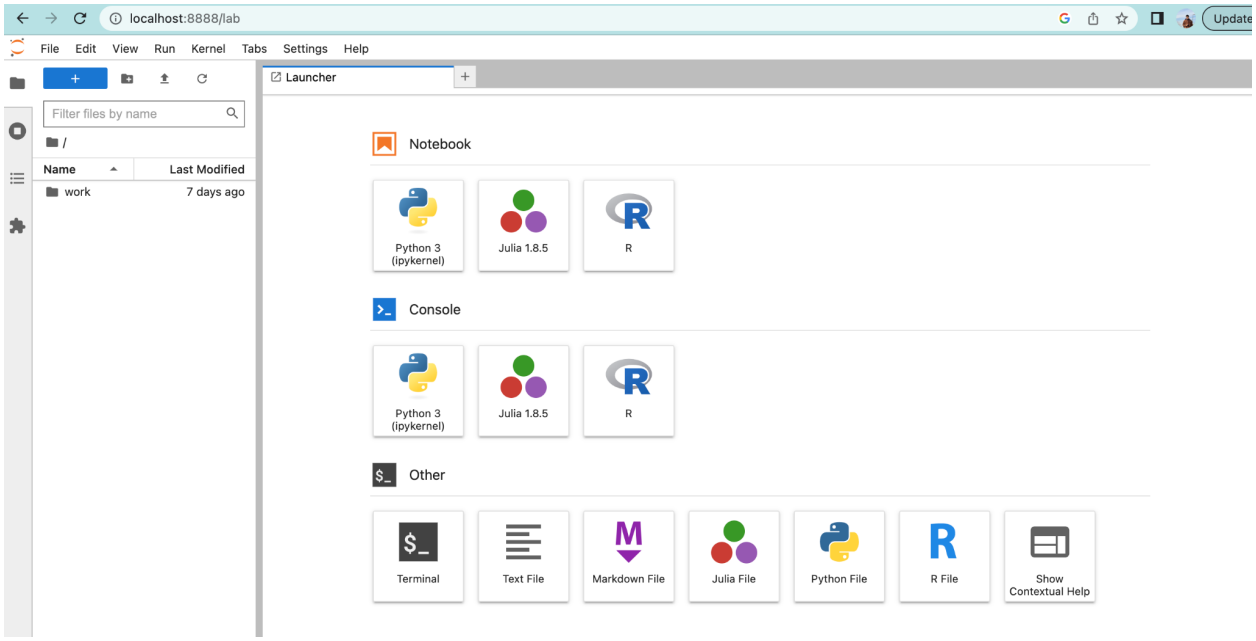
5. Expand the **Option Setting** and fill in each input as follows:
    a. **Container name:** notebook
    b. **Host port:** 8888
    c. **Container path:** /home/jovyan/work
    d. **Host path:** *<path/to/your/lab/folder>*



6. Click **Run**, then you will be routed to **Logs.** Copy the highlight token from your log console.



7. Open the browser and paste the copied highlight URL.



Now we have successfully installed JupyterLab via Docker. We can now start learning, exploring, analyzing, and visualizing data.

Python Packages

In this section, we will overview the significant and functionalities of Python packages that are commonly utilized in scientific computing, data visualization, data analysis, and machine learning.

JSON

The JSON Python package provides a functionalities to encode and decode JSON (JavaScript Object Notation) data, making it easier to work with data coming from different sources. JSON is a standard and lightweight format commonly used for data exchange between applications, especially in a web-application.
Link: https://docs.python.org/3/library/json.html

NumPy

NumPy provides support for large, multi-dimensional arrays and matrices, along with a library of mathematical functions to operate on these arrays. Numpy offers fast and efficient matrix computation. It is widely used in scientific computing, data analysis, machine learning, and deep learning applications.
Link: https://numpy.org

SciPy

SciPy provides additional functionality for scientific computing on top of NumPy. It includes modules for statistics, optimization, integration, interpolation, signal processing, linear algebra, and many other areas of mathematics and science for solving complex mathematical problems.
Link: https://scipy.org

Pandas

Pandas is fast, powerful, flexible tools designed for data manipulation and analysis. Pandas provides data structures, such as data frames, and tools for working with structured data.
Link: https://pandas.pydata.org

Scikit-Learn

Scikit-learn (Sklearn) provides simple and efficient tools for predictive data analysis tools for data mining, data analysis, and predictive modeling. It includes modules for data preprocessing, regression, classification, clustering, dimensionality reduction, and many other machine learning algorithms.
Link: https://scikit-learn.org

Matplotlib

Matplotlib is a data visualization tool providing a comprehensive set of tools for creating publication-quality figures and graphs in the form of static, animated, or interactive visualizations. It can be used to create a wide range of visualizations, such as line plots, scatter plots, bar plots, histograms, and many others.
Link: https://matplotlib.org

Seaborn

Seaborn is a tool for data visualization. Built on top of Matplotlib, it provides a higher-level interface for creating informative and attractive statistical graphics, making it easier to explore and understand data. Similar to Matplotlib, Seaborn also provides a wide range of visualization types, such as heat maps, scatter plots, bar plots, and many others.
Link: https://seaborn.pydata.org