JupyterHub Setup Guide

Advanced Data Analytics for Management Support

Chair of Information Systems, Humboldt University Berlin

Updated: March 31, 2023

Introduction

The tutorial materials typically consist of Jupyter notebooks. We will make these materials available via our GitHub repository.

You are encouraged to familiarize yourself with the concepts of GitHub and Jupyter notebooks. Introductory materials are available on the course Moodle page as well as on the web and on YouTube.

The recommended way to run Jupyter notebooks is to use the remote environments provided by the Humboldt-University of Berlin on JupyterHub. The following slides explain the steps necessary to set up this service, in depth.

Other cloud options include Google Colab and Amazon AWS. If you are comfortable installing software on your private computer and are willing to do some configuration work you can also execute Jupyter notebooks locally on your computer using JupyterLab or Visual Studio Code. This software is free to use and you do not need an account with Google or any other cloud provider. Note however, that the chair will not be able to provide technical support if you choose a different option than JupyterHub.

Step 1: Download the ADAMS repository to your computer

Install Git and optionally GitHub Desktop

https://git-scm.com/
https://desktop.github.com/ (optional)

Clone the GitHub repository

Open the GitHub Desktop App and clone the repository from our URL ...

... or type into the terminal:

```
git clone https://github.com/Humboldt-WI/adams
```

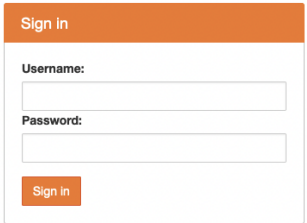
Step 2: Set up the HU VPN and visit JupyterHub

Set up the HU VPN following the instructions from the CMS

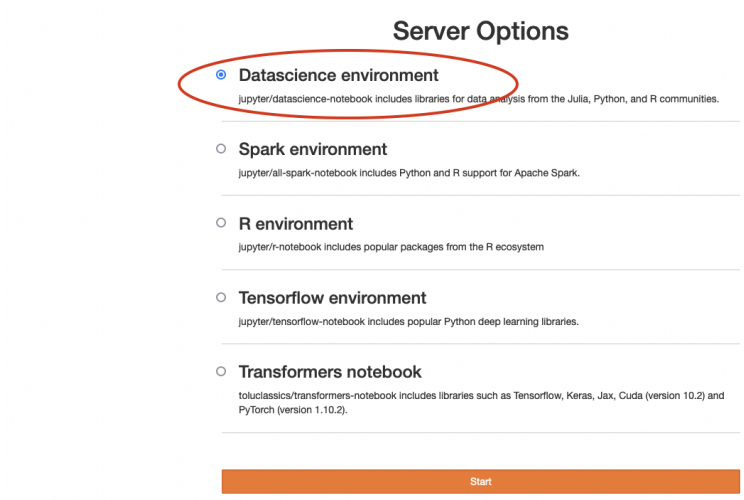
https://www.cms.hu-berlin.de/en/dl-en/netze-en/vpn/ssl-vpn/ssl-vpn

Visit the HU JupyterHub and log in with your HU credentials

1. Connect to the VPN and open JupyterHub: https://jupyterhub.cms.hu-berlin.de/
2. Type in your username and password as used in Agnes into the login form:



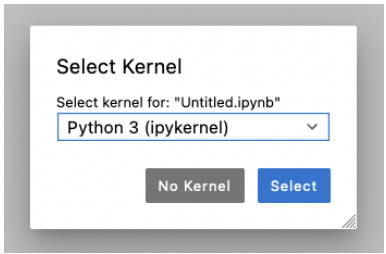
Step 3: Start the “Datascience environment”



Step 4: Select kernel

Select Python 3 kernel

You should now be greeted by a classic JupyterLab environment and prompted automatically to select a kernel. We will be using Python 3, so please go ahead and select “Python 3 (ipykernel)”. If you’re not prompted, just click on the Python 3 thumbnail in the section “notebook”.



Step 5: Upload requirements file

Locate requirements file on your computer

The ADAMS repository contains a “requirements.txt” file. You can also find it directly under this link.

If you have cloned the repository to your computer you will be able to find this file in your file explorer or through GitHub Desktop.

Upload requirements file to JupyterHub

Click on the upload symbol in the top left of JupyterHub: 

Then select the “requirements.txt” file from the ADAMS repository and upload it.

Step 6: Install requirements and get started

Install the required packages using pip

Click on the empty code cell in the new notebook that opened up (“Untitled.ipynb”) and type:

```
!pip install -r requirements.txt
```

Then press Shift+Enter to run the code cell. All required packages will automatically install. This should take less than a minute.

Unfortunately, due to the nature of JupyterHub you will need to reinstall the requirements everytime you reconnect to the server. That is a drawback that all cloud-based solutions share.

Upload Jupyter notebook or start coding

You’re done setting up the environment! You can now upload Jupyter notebooks from the ADAMS repo to JupyterHub using the upload button  or get started coding directly in the Juypter notebook that you’re in.