**Instructions for executing the DASK Demo**

**Video Link (Youtube) – Demo for showing how to execute DASKDemo**

<https://youtu.be/x4HGbh8mJNQ>

**Installation Instructions for DASK**

**Installed in Windows using Anaconda**

conda install dask

conda install dask distributed

**For displaying graphs that we can schedule using visualize function of dask**

conda install graphviz

**For installing the WEB UI – Bokeh you can install it by executing the following commands**

**PIP**

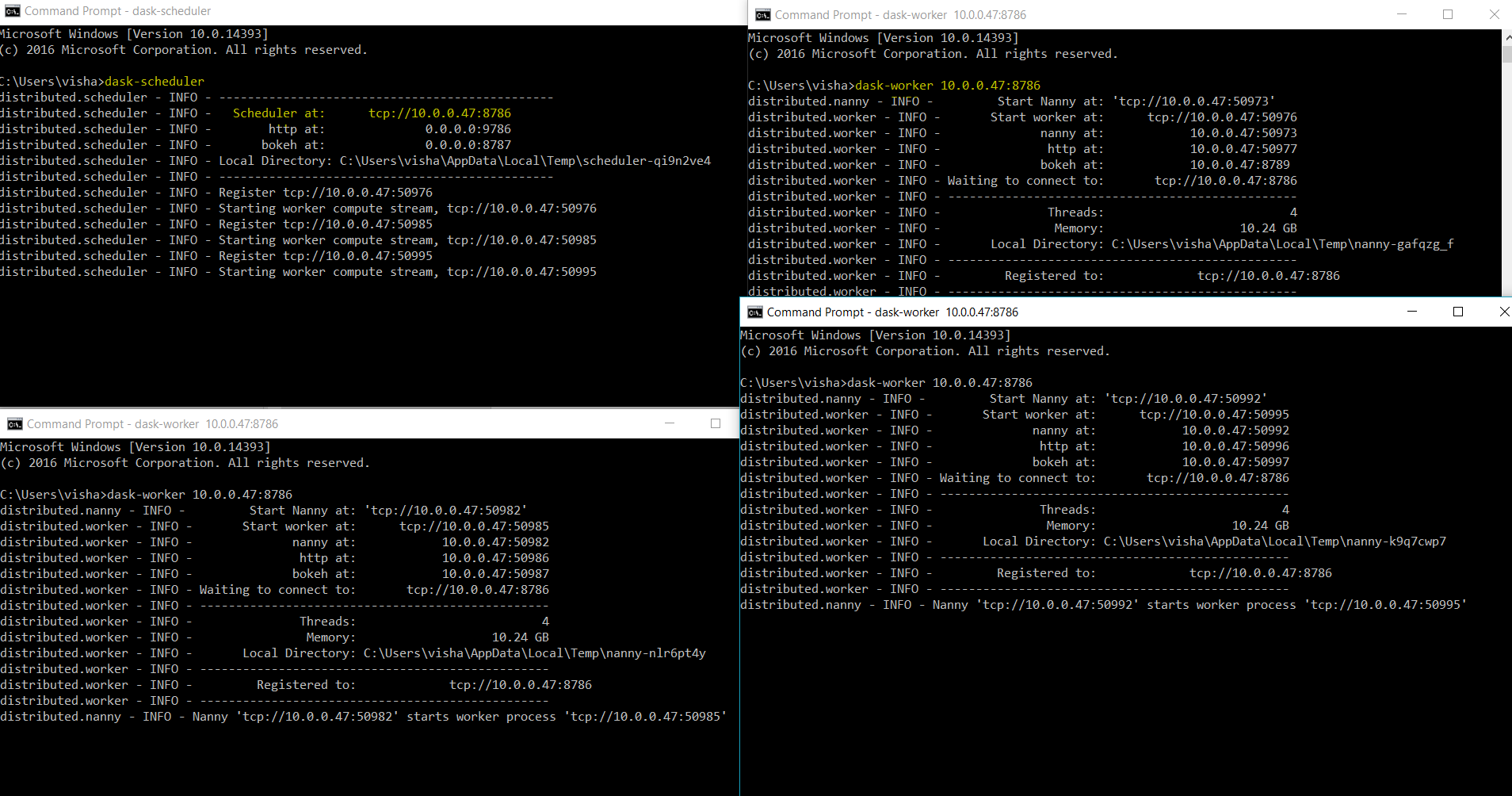
pip3 install bokeh

pip install bokeh

**Anaconda**

conda install bokeh

**Create a local Dask distributed environment by invoking dask-scheduler daemon in one terminal and connecting to the ip address and port of this using dask-worker daemon threads running in different terminals.**



**Executing IPython Notebook**

You can try executing the code DaskDemo.ipynb IPython notepbook which you can download from the DASKDemo folder on github

You might face some issue with module dependencies. In that case, Please notify us of the same and we will add them in this document.

**For running DASK on AWS.**

1. You will need to have AWS credentials.
2. Add these credentials as **Environment variables** ( **AWS\_ACCESS\_KEY\_ID and AWS\_SECRET\_ACCESS\_KEY** )
3. Install the dask-ec2 package from pip using pip install dask-ec2
4. By default, the EC2 instances are created on the us-east-1 region. So, you should create a KeyPair on the the us-east-1 (N.Virginia) region.
5. Save the generated dasktest.pem file (private key) which AWS provides and provide the path of that file in the above command. In our case, we execute this command from the directory in which dasktest.pem file is located.
6. You will need SSH to be installed in your system to be able to execute these commands. We recommend executing the above command using Cygwin terminal available for Windows after checking if you have ssh. (You can verify if you have ssh by executing ssh <some url> and see the output on the console).

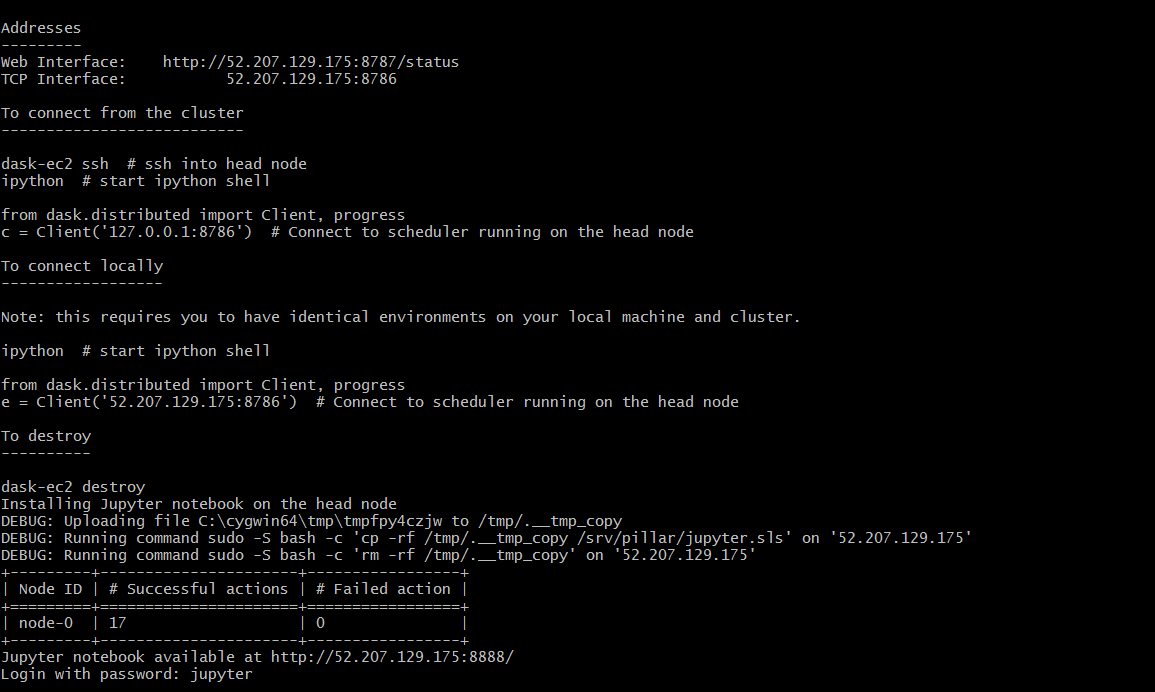
Mac users only need to make sure that they have ssh setup in their system.

1. Execute the command

dask-ec2 up --keyname dasktest --keypair dasktest.pem --count 3 --nprocs 3 --tags v:vish



1. On executing the above command dask-ec2 will configure and install 3 (specified by count) EC2 instances on AWS.
2. Once AWS cluster starts up, it will return the jupyter notebook URL which you can use to run your analysis on AWS.



1. AWS also returns the URL of the dask-scheduler which you can use to open the Bokeh Web UI to visualize the parallel execution of your tasks. Bokeh is usually available at the 8787 port
2. You can use the code from the AWSDASKDemo.ipynb file provided in the DASKDemo folder from github.