1. Install h5py

On Ubuntu. You can try the following three commands:

1. sudo pip install cython
2. sudo apt-get install libhdf5-dev
3. sudo pip install h5py

source: <https://github.com/fchollet/keras/issues/3426>

1. install python3.6
   1. sudo add-apt-repository ppa:jonathonf/python-3.6
   2. sudo apt-get update
   3. sudo apt-get install python3.6

sudo apt-get install software-properties-common

sudo add-apt-repository ppa:deadsnakes/ppa

sudo apt-get update

sudo apt-get install python3.6

1. update python3/python to new version python3.6

sudo update-alternatives --install /usr/bin/python python /usr/bin/python2.7 1

sudo update-alternatives --install /usr/bin/python python /usr/bin/python3.6 2

sudo update-alternatives --config python

1. install pip

sudo apt install python3-pip

sudo pip install --upgrade pip

1. install h5pyd

pip install --user git+http://github.com/HDFGroup/h5pyd.git

vi ~/.hscfg

# HDFCloud configuration file

hs\_endpoint = https://developer.nrel.gov/api/hsds/

hs\_username = None

hs\_password = None

hs\_api\_key = 3K3JQbjZmWctY0xmIfSYvYgtIcM3CN0cb1Y2w9bf

1. Read h5 file on s3 buckets

1. local: 353ms, 222ms(extraction)

2. s3/h5py: 2.77s, 10.3s(extraction)

3. HSDS server: 1.6s (can catch, decrease to 834ms), 103m

1. install s3fs

pip install awscli --upgrade –user

pip install boto3

pip install s3fs

1. Install airflow

export AIRFLOW\_HOME=~/airflow

sudo apt install python3.6-dev

pip install apache-airflow

airflow initdb

airflow webserver -p 5000

airflow scheduler