**Installation of Airflow**

**windows or linux ?  
  
cloudera quickstart VM  
  
it has linux - centos  
  
Airflow  
========  
  
1. create workflows  
  
2. schedule workflows  
  
3. monitor workflows  
  
  
In cloudera quickstart vm we have linux centos 6.7 installed.  
  
  
we need to have python installed for airflow to run.  
  
we need a python version above 3.5  
  
what version of python is there in cloudera quickstart VM?  
  
Python 2.6.6  
  
if we change this base version to 3.5 or above, some of the services might mis behave.  
  
we do not want to change the base version that is 2.6.6  
  
simplest way to do this..  
  
1. Install the SCL utility.  
  
software collections.  
  
its goal is to let you install and run multiple versions of the same software on the same system. we dont want to interfere with the sytem version of python.  
  
sudo yum install centos-release-scl  
  
  
2. we will install python 3 with using SCL  
  
sudo yum install rh-python36  
  
  
3. check the python version  
  
python --version (2.6.6)  
  
  
4. Launch newer version of python using SCL shell  
  
scl enable rh-python36 bash  
  
5. check the python version again  
  
3.6.9  
  
  
6. create our own project folder  
  
mkdir airflow  
  
cd airflow  
  
  
7. create a python virtual environment  
  
python -m venv venv  
  
this command will create a venv folder inside airflow folder.  
  
8. activate this virtual environment which we created just now.  
  
when running the below command you should be inside airflow  
source venv/bin/activate  
  
9. install apache airflow.  
  
pip install apache-airflow  
  
airflow is now installed and that too successfuly.  
  
airflow  
  
  
10. airflow requires a sql like database.  
  
airflow initdb  
  
11. set the airflow path  
  
export AIRFLOW\_HOME=$PWD  
  
echo $AIRFLOW\_HOME (VERIFICATION)  
  
12. open one more terminal  
  
and activate the virtual environment  
  
cd airflow  
  
source  venv/bin/activate  
  
13. start the webserver in one terminal and  
start the scheduler in other terminal  
  
airflow webserver -p 8080 (terminal 1)  
  
airflow scheduler (terminal 2)**

Airflow setup

We are providing a VM which has airflow installedcalled a clone VM. Justdownload , import the VM and start using it as steps given below

<https://drive.google.com/drive/folders/1fXViofZOqdvX2bKY3RVwyTpCbgvFS7iI?usp=sharing>

[If you are unable to import the VM and if you seea .vdi instead of .ovf afterextracting, follow the steps below to import the VMto Virtual boxEither right click on the .vdi file and open withVirtual box

OR

Click on the+ADD option in the Virtual Box and selectthe .vdi file. This should start the import with default values.]

Follow below steps to setup airflow1. su2. cd airflow3. source venv/bin/activate4. export AIRFLOW\_HOME=$PWD5. Create dag folder if not exists6. Create python file inside dag folder7. Airflow web server -p 80808. Open another terminal — do steps 1-4 — Airflowscheduler

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated