Airflow Instructions

Step 0 Airflow Installation

```
$ pip install apache-airflow
```

Step 1: Airflow Initialization

a

In terminal, navigate to your favorite place and make a directory as the workspace for airflow.

```
$ cd /../Documents
$ mkdir airflow_tutorial
```

b

Set the AIRFLOW_HOME environment to this directory.

```
$ export AIRFLOW_HOME=`pwd`/airflow_tutorial
```

c

Type the following two commands to initialize airflow in the folder.

```
$ airflow version
$ airflow initdb
```

The first command will generate file **airflow.cfg** and **unittests.cfg**; The second command will generate file **airflow.db**.

d

Start the Airflow webserver

```
$ airflow version
$ airflow initdb
```

After this configuration, open your web browser and go to http://localhost:8080/admin/, you can see the web server of airflow with a bunch of default DAGs listed.

Step 2: DAG – Directed Acyclic Graph

a

While keep the previous terminal window running, open a new terminal window and navigate to the work space and set the AIRFLOW_HOME environment again.

```
$ cd /../Documents
$ export AIRFLOW_HOME=`pwd`/airflow_tutorial
```

b

Navigate to the folder airflow_test, and make another directory called 'dags'.

```
$ cd airflow_test
$ mkdir dags
```

c

Open file **airflow.cfg** and make sure 'dags_folder' is set to the folder you just created. Otherwise, change accordingly using the follow code.

```
dags_folder = /../Documents/airflow_tutorial/dags
```

d

In the dags folder, add the following code and save it as a python file called 'hellow_world.py'

Step 3: Push the DAG to the webserver.

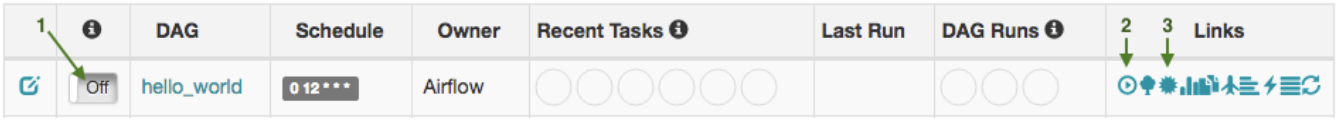
a

Type the following code to push the dag to the webserver.

```
airflow scheduler
```

b

Refresh the webserver, and you should be seeing the following DAG that is just added.



c

Click 1 to turn the DAG on, you might need to refresh the page after this. Then click 2 to run the DAG. Finally, you can click 3 to see the DAG and its status.

A new DAG to build a pipeline

Now you just made airflow work on your machine, it's time to build up the first practical airflow pipeline.

a

Navigate to the airflow workspace you just created and make a new directory called 'tasks', where we will save three python files. Our goal is to use airflow to build a pipeline so that these three files will run automatically and consecutively.

```
$ cd /../Documents/airflow_tutorial
$ mkdir tasks
```

Save these three python files in this directory. These files will each write out a csv file as a result.

b

In the same 'dags' folder, create another file called 'csv_writer.py', and paste in the code from here.

c

Then we just need to follow the same procedure as above to make it run. In a terminal window, run:

```
$ cd /../Documents
$ export AIRFLOW_HOME=`pwd`/airflow_tutorial
$ airflow webserver
```

Start a new terminal window, and do:

```
$ cd /../Documents
$ export AIRFLOW_HOME=`pwd`/airflow_tutorial
$ airflow scheduler
```

d

Last, repeat Step 3(b) and 3(c) to finish running it in the webserver.