



Installing Airflow

How to install Apache Airflow



Start Your VM

1. When your VM is running you can either:
 - a. Use your terminal by connecting it through ssh with the command:
 - i. `ssh -p 2200 airflow@localhost`
 - b. If you are on Windows, you can use Putty: [How to use SSH in Putty](#)
2. Then execute the following commands:
 - a. `su -` (then enter the root password: centos)
 - b. `start-dfs.sh`
 - c. `exit`
 - d. `hive --service metastore &`
 - e. `sudo /opt/presto/bin/launcher start` (then enter the airflow password: airflow)
 - f. `sudo /opt/spark/sbin/start-master.sh`
 - g. `sudo /opt/spark/sbin/start-slaves.sh`



Installing Airflow

Type the following commands:

1. `source .sandbox/bin/activate`
2. `pip install "apache-airflow[celery, crypto, postgre, hive, rabbitmq, redis]"`
 - a. Because of a GPL dependency (licence), you will get an error. Just run:
`AIRFLOW_GPL_UNIDECODE=yes pip install apache-airflow`
3. `cd ~`
4. `airflow initdb`
5. `airflow upgradedb`
6. `mkdir -p ~/airflow/dags`



Installing Airflow

In the `~/airflow/airflow.cfg` file change the following:

| Parameter | Old Value | New Value |
|------------------------------|------------------------------------|------------------------------------|
| <code>base_url</code> | <code>http://localhost:8080</code> | <code>http://localhost:8083</code> |
| <code>web_server_port</code> | <code>8080</code> | <code>8083</code> |
| <code>load_examples</code> | <code>True</code> | <code>False</code> |



Installing Airflow

1. `cd ~`
2. `airflow resetdb` (to get rid of DAG examples)
3. `cp airflow_files/hello_world.py airflow/dags/`



Installing Airflow

- Run the web UI:
 - `airflow webserver`
- Open a new terminal and type:
 - `ssh -p 2200 airflow@localhost`
 - `source .sandbox/bin/activate`
 - `airflow scheduler`
- Now go to your browser and open <http://localhost:8083>

What Do We Have?

```
[airflow@localhost ~]$ tree airflow/
```

```
airflow/
```

```
├── airflow.cfg
```

Fichier de configuration

```
├── airflow.db
```

Database (SQLite)

```
├── airflow-webserver.pid
```

Pid of the web server

```
├── dags
```

Folder where you will put your DAGs

```
├── logs
```

Log folder

```
└── scheduler
```

```
    ├── 2018-10-08
```

```
    └── latest -> /home/airflow/airflow/logs/scheduler/2018-10-08
```

```
└── unittests.cfg
```

Unit tests configuration

```
5 directories, 4 files
```