Airflow 中最基本的兩個概念是：DAG 和 task。

DAG 的全稱是 Directed Acyclic Graph 是所有你想執行的任務的集合，

在這個集合中你定義了他們的依賴關係，一個 DAG 是指一個 DAG object，一個 DAG object 可以在 Python 腳本中配置完成。

比如一個簡單的的 DAG 包含三個 task：A、B、C，

A 執行成功之後 B 才能執行，C 不依賴 A 和 B 即可執行。

在這個簡單的 DAG 中 A B C 可以是任何你想要執行的任務。

DAG 的定義使用 Python 完成的，其實就是一個 Python 檔，存放在 DAG 目錄，Airflow 會動態的從這個目錄構建 DAG object，每個 DAG object 代表了一個 workflow，每個 workflow 都可以包含任意多個 task。

安裝和使用Airflow是基於Python構建的，可以很容易用pip安裝使用，pip install apache-airflow，默認情況下airflow會在~/airflow目錄存放相關配置。

Airflow 提供了一些列命令來完成airflow 的初始化工作來和它的正確使用。

# 在 airflow 目錄

初始化資料庫和 airflow 配置

airflow initdb

pip install "apache-airflow[crypto, slack]"

export AIRFLOW\_HOME="$(pwd)"

airflow initdb

理想上把 AIRFLOW\_HOME 加入到 ~/.bash\_profile 裡頭之後會比較輕鬆

---

The example files are not in /usr/local/airflow/dags. You can simply mute them by edit airflow.cfg (usually in ~/airflow). set load\_examples = False in 'core' section.

---

airflow.exceptions.AirflowException: Could not create Fernet object: Incorrect padding

出現這個問題的原因在於fernet key 需要修改airflow.cfg (默認位於~/airflow/)裡的fernet\_key

使用python -c "from cryptography.fernet import Fernet; print(Fernet.generate\_key().decode())"這個命令生成一個key

然後替換airflow.cfg裡的fernet\_key

再執行airflow initdb或airflow resetdb 就可以了

---

輕鬆排程

with DAG('comic\_app\_v1', default\_args=default\_args) as dag:

...

這行實際上就定義了我們的 DAG 並將它命名為 comic\_app\_v1。而此 DAG 的排程（Scheduling）設定如

* 'start\_date': datetime(2100, 1, 1, 0, 0) 代表從西元 2100 年開始第一次執行此 DAG
* 每次執行之間間隔多久。'schedule\_interval': '@daily' 代表每天執行一次
* 'retries': 2 則允許 Airflow 在 DAG 失敗時重試 2 次
* DAG 失敗後等多久後開始重試（'retry\_delay': timedelta(minutes=1)　代表等一分鐘）
* 更多更多 ...

這些排程設定為了方便管理，一般都另外定義在 default\_args 變數並放在 script 的最上面

在 Airflow 裡頭，DAG 只知道有哪些工作以及這些工作之間的執行順序。而實際上這些工作要怎麼被完成，其實作邏輯則是由各種 [Operator](https://airflow.apache.org/code.html#operators) 負責。

你可以想像 [Opeartors](https://airflow.apache.org/code.html" \l "airflow.operators) 就是幫我們完成特定種類工作的小幫手，像是一些常見的例子：

* [PythonOperator](https://airflow.apache.org/code.html#airflow.operators.PythonOperator) 執行一個 Python 函式
* [BashOperator](https://airflow.apache.org/code.html#airflow.operators.BashOperator) 執行 Bash 指令
* [S3KeySensor](https://airflow.apache.org/code.html#airflow.operators.S3KeySensor) 監測 S3 上的檔案存不存在
* [SlackAPIPostOperator](https://airflow.apache.org/code.html#airflow.operators.SlackAPIPostOperator) 送訊息給 Slack
* ...