目標：建立一套自動爬蟲系統

方式：CloudWatch設定時間條件，觸發Lambda執行函式，對EC2執行個體做開關機的動作，並利用Crontab設定執行個體內執行排程，將爬取的資料存入RDS資料庫

使用AWS服務：EC2、Lambda、CloudWatch、IAM、RDS

**步驟一、建立RDS資料庫**

1. 開啟RDS服務
2. 點擊左方列表"資料庫"
3. 點選右方按鈕"建立資料庫"
4. 資料庫建立方法選擇"標準"
5. 引擎我這邊選擇"MySQL"
6. 設定自己的"使用者名稱"及"主要密碼"，待會連線到MySQL會需要
7. "連線"區塊的"公開存取"務必選擇"是"
8. 接下來就點選"建立資料庫"，資料庫就建立完成
9. 可以透過MySQL的[WorkBench](https://dev.mysql.com/downloads/workbench/)連線到資料庫內

**步驟二、建立EC2執行個體**

1. 開啟EC2服務
2. 點選左列"執行個體"
3. 點擊右方按鈕"啟動新執行個體"，這邊我創建的是一台Ubuntu的Server
4. "金鑰對 (登入)"欄位務必建立新的金鑰對並妥善保存.pem檔案
5. 網路部分勾選"允許 SSH 流量" 來自0.0.0.0/0
6. 儲存空間免費帳號可以使用到30GB
7. 點選"啟動執行個體"完成建立
8. 可透過SSH連線到執行個體內，並透過scp指令將爬蟲檔案傳入虛擬機內，再使用crontab -e指令編輯cron的設定檔設定自動執行排程

**步驟三、建立IAM角色**

1. 開啟IAM服務
2. 點選左列"角色"
3. 點選右方"建立角色"
4. 使用案例選擇"Lambda"並點擊"下一步"
5. 進入"選擇政策"頁面，點擊右方"建立政策"
6. 在建立政策頁面點擊"JSON"欄位貼入以下程式碼

{

"Version": "2012-10-17",

"Statement": [

{

"Effect": "Allow",

"Action": [

"logs:CreateLogGroup",

"logs:CreateLogStream",

"logs:PutLogEvents"

],

"Resource": "arn:aws:logs:\*:\*:\*"

},

{

"Effect": "Allow",

"Action": [

"ec2:Start\*",

"ec2:Stop\*"

],

"Resource": "\*"

}

]

}

* 7.檢閱後輸入名稱並建立
* 8.回到"選擇政策"頁面，選擇剛剛建立的政策並完成建立角色

**步驟四、建立關機Lambda函式**

1. 開啟Lambda服務
2. 點選左方"函數"，並點擊右方"建立函式"按鈕
3. 選擇"從頭開始撰寫"
4. 取一個可以識別函式的名稱，執行時間選擇 Python 3.7
5. "**變更預設執行角色**"區塊選擇**步驟三**建立的角色，就可以點選"**建立函式**"
6. 建立完成後，往下可看到Lambda函式區塊，貼上以下程式碼後點選Deploy

import boto3

region = 'us-west-1' →改成你使用的區域

instances = ['i-12345cb6de4f78g9h', 'i-08ce9b2d7eccf6d26']　→設定你想關機的EC2執行個體ID，可以到EC2執行個體介面找到ID，需使用陣列定義

ec2 = boto3.client('ec2', region\_name=region)

def lambda\_handler(event, context):

ec2.stop\_instances(InstanceIds=instances)

print('stopped your instances: ' + str(instances))

**步驟五、建立CloudWatch事件**

1. 在步驟四建立的"函式概觀"內，點選"新增觸發"
2. 選取"EventBridge(CloudWatch事件)"
3. 規則區塊點選"建立新規則"，取一個可識別的名稱
4. 選擇排程表達式，並依照你希望開關機的時間做設置，Ex：(30 2 \* \* ? \*)，這邊時間要注意是UTC+0的時區
5. 完成即可建立觸發條件

**步驟六、建立開機Lambda函式**

1. 依照步驟四及步驟五順序建立開機函式
2. 需要注意的是程式碼中ec2.stop\_instances(InstanceIds=instances)這行需改為ec2.start\_instances(InstanceIds=instances)
3. CloudWatch開機時間需略早於步驟二中設定的執行時間，讓虛擬機有時間可以開啟

操作完成以上步驟即可擁有一套每日自動更新的資料庫：）