Contents

Introduction

This is a research report that generalizes the steps to automate a script to run at any given time so that the data in csv can be pushed into phpmyadmin. This report assumes the fact that you are able to access a server internally and also able to access the database externally.

Steps to follow

Installation of linux

You can skip this step if you have a linux server installed.

Direct access to server

This option is available IF you have the credentials of the server.

Prerequisites

Before you can establish a secure remote desktop protocol with a remote machine, there are a few basic requirements to meet:

The remote computer must be turned on at all times and have a network connection.

The client and server applications need to be installed and enabled.

You need the IP address or the name of the remote machine you want to connect to.

You need to have the necessary permissions to access the remote computer.

Firewall settings need to allow the remote connection.

Command

Run command prompt / terminal and enter the following command

ssh *host

*host refers to the ip address of the server you are intending to automate the script on. Eg. 111.222.1.2

```
selvania@111.222.1.2's password:
```

Next, you will have to enter the password for the server.

If it's the first time that you are accessing the server, this warning will pop out

```
The authenticity of host 'localhost (127.0.0.1)' can't be
established. ECDSA key fingerprint is
SHA256:9jqmhko9Yo1EQAS1QeNy9xKceHFG5F8W6kp7EX9U3Rs. Are you sure you
want to continue connecting (yes/no)?
```

Type in yes and press enter

```
Warning: Permanently added 'localhost' (ECDSA) to the list of known
hosts.
```

This will show and the host should be added into the known hosts.

If there are no error you should be inside the server.

Creation of an ubuntu server

There are mainly 3 ways of creation. First way is to download the ubuntu iso and install it locally via virtualbox (not recommended due to how firewall and connection works). Second way is to spin up a server on either Amazon web service /Google cloud platform or microsoft azure. This method will require payment so be aware. Lastly is to install ubuntu on your computer.

Once you have spin up the server, refer back to Direct access to server for information to connect to the server.

Pushing the scripts

If you reached this step, it means that you are able to successfully enter the server. Congratulations
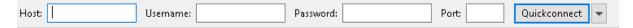
The next step is to push the scripts into the server for usage.

I recommend this software for this job.

Filezilla

You are free to choose other softwares as long as it gets the job done.

Enter in the host, user and password. Port can remain empty or will be 21 for ftp, 22 for sftp

| Host: | | Username: | | Password: | | Port: | | Quickconnect | ▼ |

Once connected, you will be able to transfer files into the server.

Transfer the 3 scripts into the server and this step is completed.

Running the scripts

Run the scripts on the server to test whether it works on that environment. If it doesn't, troubleshoot until you are able to get the csv files that you wanted. Everything should be self-explanatory here.

Automation

Via crontab

This option is assuming that you have to run the script hourly/daily and you do not want to go in hourly/daily to run it manually.

Use the command

`crontab –e`

Add lines below to run the job hourly

```
00 * * * * python3 CatalogueScraper.py
```

```
00 * * * * python3 PageScraper.py
```

```
00 * * * * python3 ReviewScraper.py
```

Save and exit. The scripts should be run automatically every hour

Pushing data to database (phpMyAdmin)

To do this, another python script is required to push the data from server to database.

```python
import pymysql

#database connection
connection = pymysql.connect(host="localhost", user="root", passwd="",
database="databaseName")
cursor = connection.cursor()
```

Fill in the relevant details from your phpmyadmin. Host will be the ip address, user and passwd are from phpmyadmin and database is the name of your database. If you can run the line above, it means that the connection is established and you can start to import the data into phpmyadmin.

```python
filenames = ['a.csv', 'b.csv', 'c.csv']
for names in filenames:
    #read from each csv file
    #get the line for each csv file
    # do an insert and execute statement for each line
    insertStatement = "insert into table (column1, column2) values('a', 'b');"
    cursor.execute(insertStatement)

# once all statements are processed. commit and close the connection
connection.commit()
connection.close()
```

The following pseudocode will allow the csv files to get pushed into the database. Save the file as ImportData.py and follow Via crontab to add another line to the list

```
30 * * * * python3 ImportData.py
```

We set it 30 minutes after all the scripts are finished so that the scraping can be done before we import. Save the file and it should be done.

End of Document

If you have finished the document, you should be able to automate the operation without doing it manually.