Ambari Server

Using the Horton Works Virtual Machine, we downloaded HortonWorks Sandbox from the official web site, but I couldn't work in the Virtual Box in our OS (not enough Required Memory).

From this problem, we started using the Google Cloud by deploying the ambary server on a Virtual Machine with Centos6 (8 vCPU, 52 Go RAM), so to do that we created an account that give us free 300$ credits that we can use for the compute engine.

First, we started by setting up password-less SSH connections between the Ambari Server host and all other hosts in the cluster. To do that we generated a SSH public and private Keys with the RSA encryption, we copy the keys generated to the target host, we added the public key to the authorized keys, we set permission for the authorized keys.

```
ssh-keygen -t rsa
.ssh/id_rsa
.ssh/id_rsa.pub
cat id_rsa.pub >> authorized_keys
chmod 700 ~/.ssh
chmod 600 ~/.ssh/authorized_keys
```

Second, we will synchronize the clock of the cluster and the machine that runs the browser that we will see later to access to the ambari server.

```
service ntpd start
chkconfig ntpd on
```

After that, we add the host of the server by adding the local IP address and the name of domain to configure the DNS protocol, to do that we edited the **vi /etc/hosts** by adding the domain that is defined for our virtual machine.

```
10.156.0.2 hortonworks.c.cloud-service-190416.internal hortonworks
```

- **hortonworks.c.cloud-service-190416.internal :** name of the project in the google cloud
- **hortonworks :** Name of the chosen domain

By adding this domain, we need to configure the network of this host, we edit the **vi /etc/sysconfig/network** by modify the HOSTNAME to the domain chosen.

So, to make the communication between the hosts and the Ambari Server, we must open the ports that we use to get to the browser, so we disabled the **iptables** and **ip6tables**

```
service iptables stop

service ip6tables stop

chkconfig iptables off

chkconfig ip6tables off
```

We disabled the SELinux and Umask,

```
vi /etc/selinux/config

SELINUX=disabled

umask 0022
```

We used MySQL database for the Hive operation in the Ambari Server, so we check for the connector.

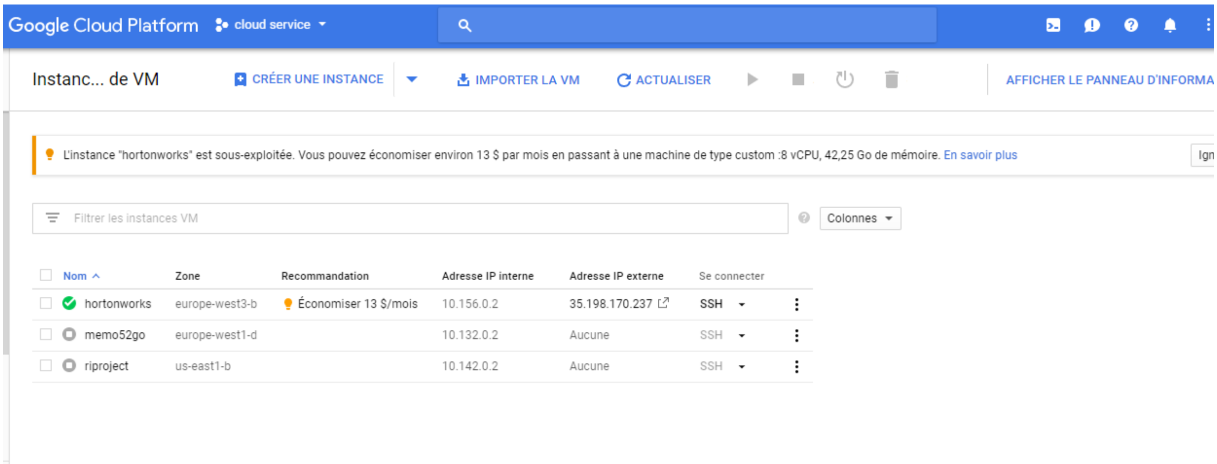
After the configuration that we made below, we downloaded the public repository Ambari Server and install it.

```
wget -nv http://public-repo1.hortonworks.com/ambari/centos6/2.x/updates/2.6.1.0/ambari.repo -O /etc/yum.repos.d/ambari.repo

yum install ambari-server

ambari-server setup

ambari-server start
```

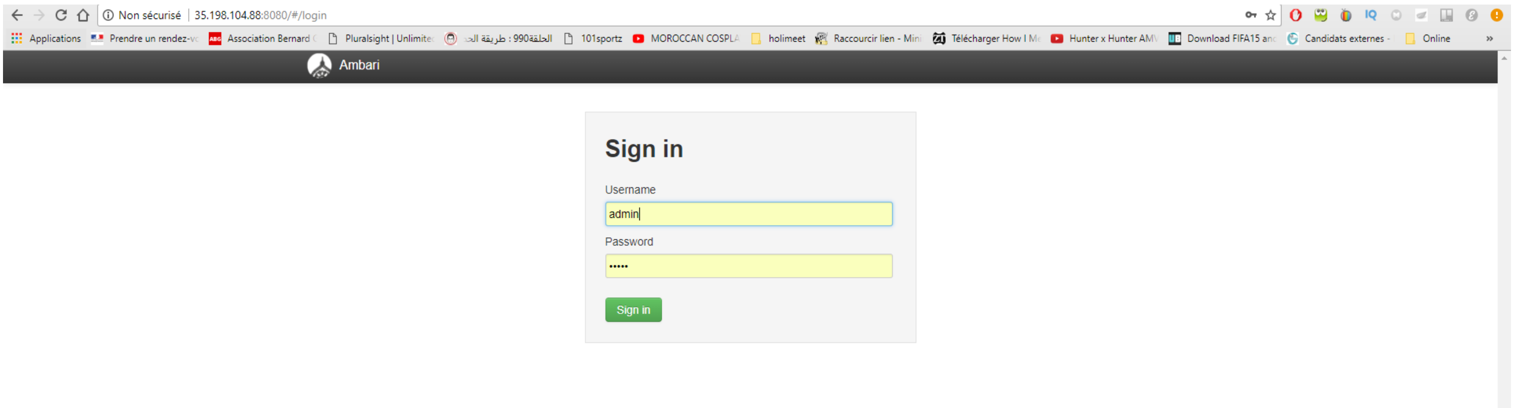


L'instance "hortonworks" est sous-exploitée. Vous pouvez économiser environ 13 $ par mois en passant à une machine de type custom :8 vCPU, 42,25 Go de mémoire. En savoir plus

□	Nom	Zone	Recommandation	Adresse IP interne	Adresse IP externe	Se connecter
□	✓ hortonworks	europe-west3-b	Économiser 13 $/mois	10.156.0.2	35.198.170.237	SSH
□	○ memo52go	europe-west1-d		10.132.0.2	Aucune	SSH
□	○ riproject	us-east1-b		10.142.0.2	Aucune	SSH

We press the SSH key.

After the installation and the running, the server is running in the external address (it changes while we restart the virtual machine) in the 8080 port,

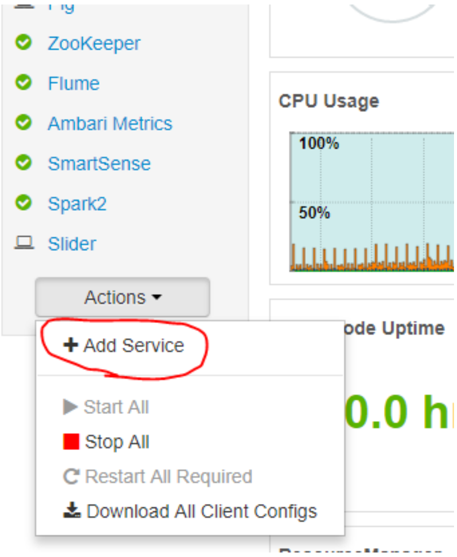


Sign in

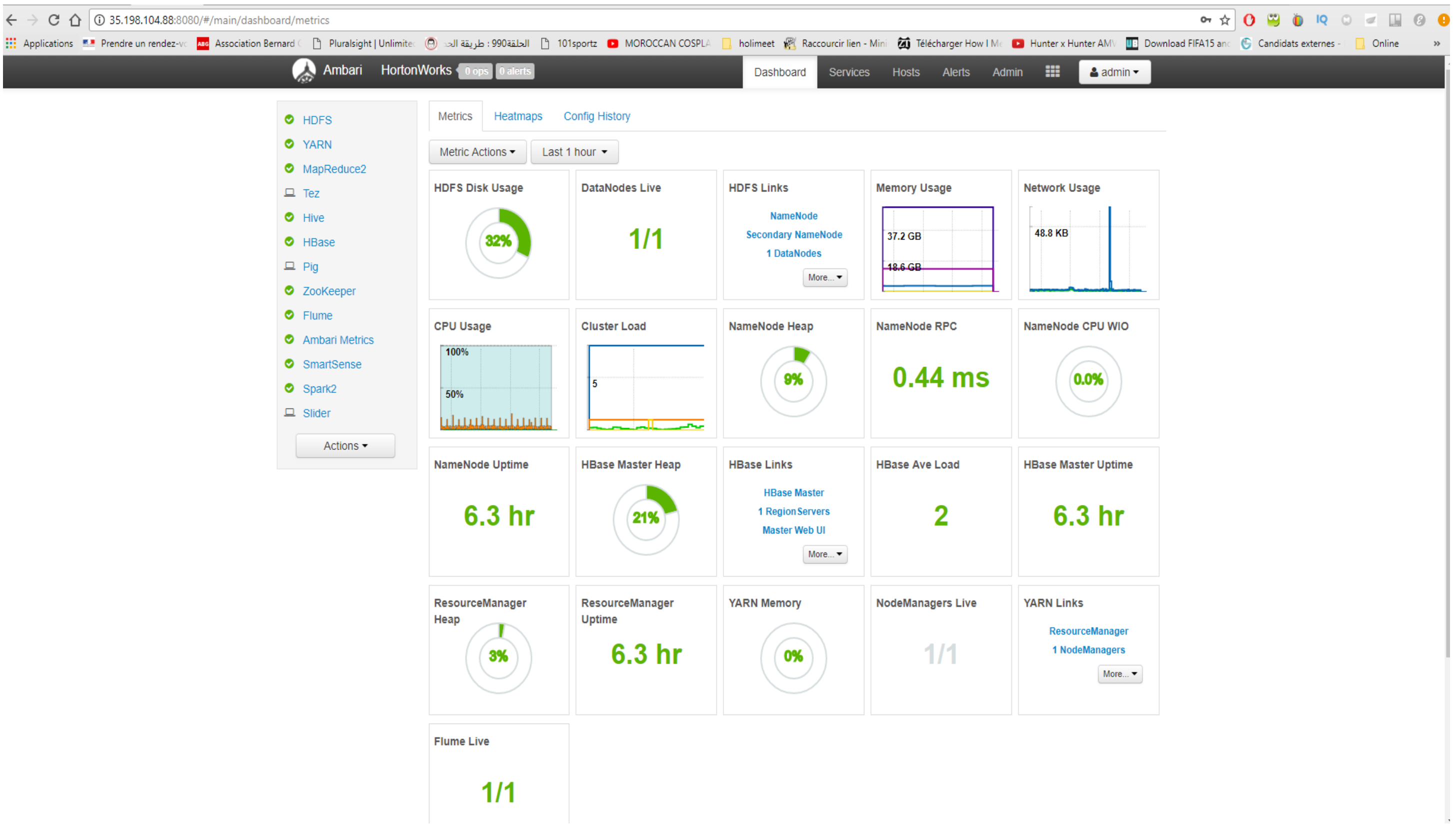
Username: admin

Password: •••••

We can log in with the **"admin" "admin"**, and right after, we configured the server by using the Private key that we generated. After, we added the services that we may use, (HDFS, MapReduce, Spark2, ZooKeeper, Hive, etc. …)



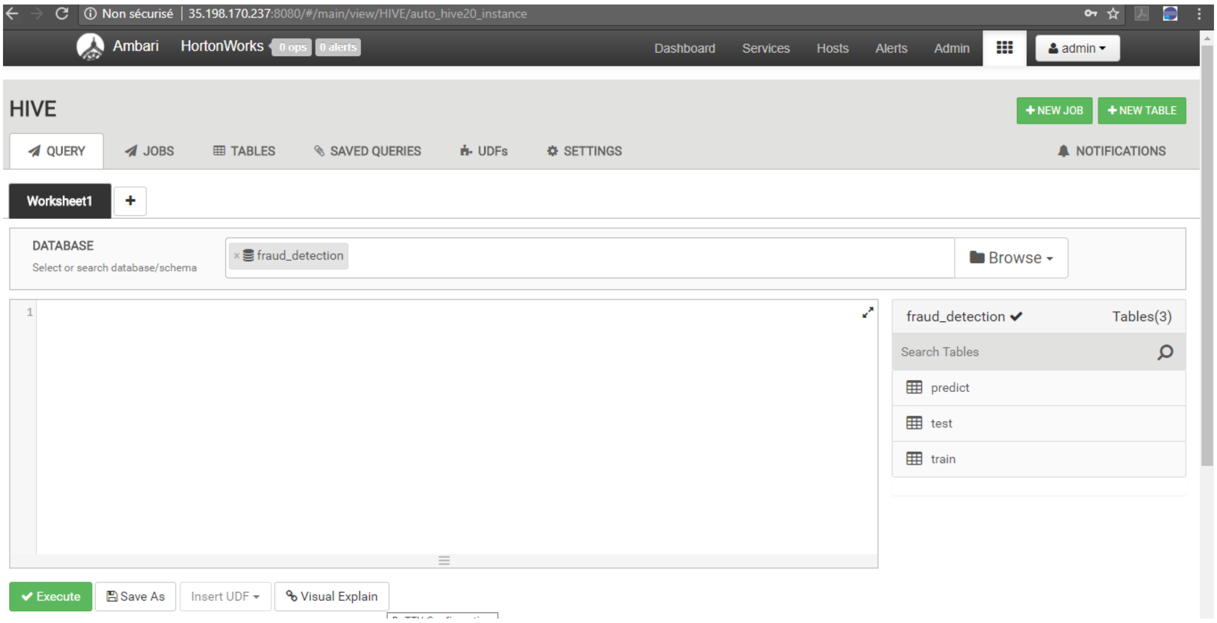
So, the dashboard looks like this:



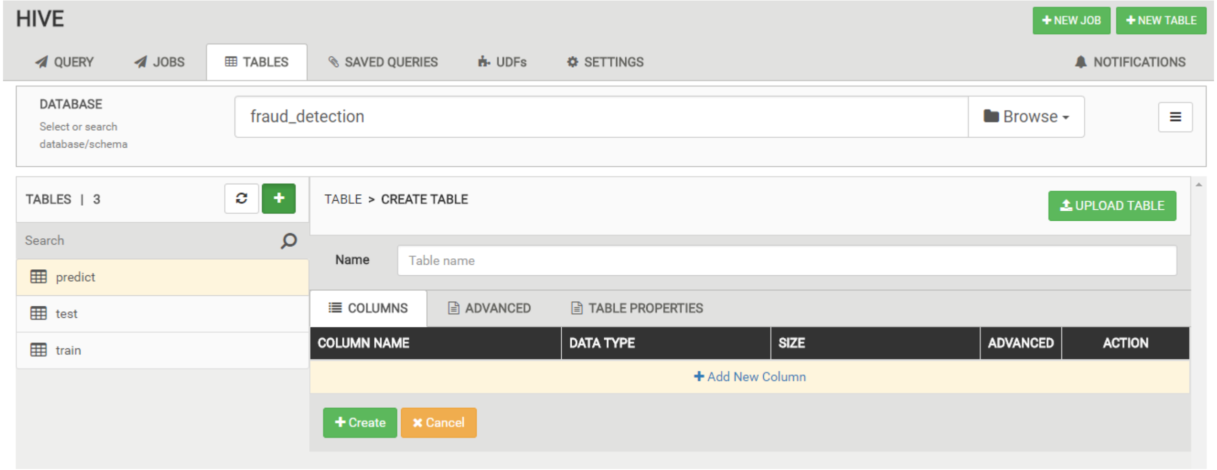
We will let the virtual machine running in the cloud for the next day, you can access to it by this link

http:// 35.198.170.237 :8080

Coming to Hive. It provides us data warehousing facilities on top of an existing Hadoop cluster. Along with that it provides an SQL like interface which makes your work easier, in case you are coming from an SQL background. You can create tables in Hive and store data there. Along with that you can even map your existing HBase tables to Hive and operate on them.



We click on New Table:

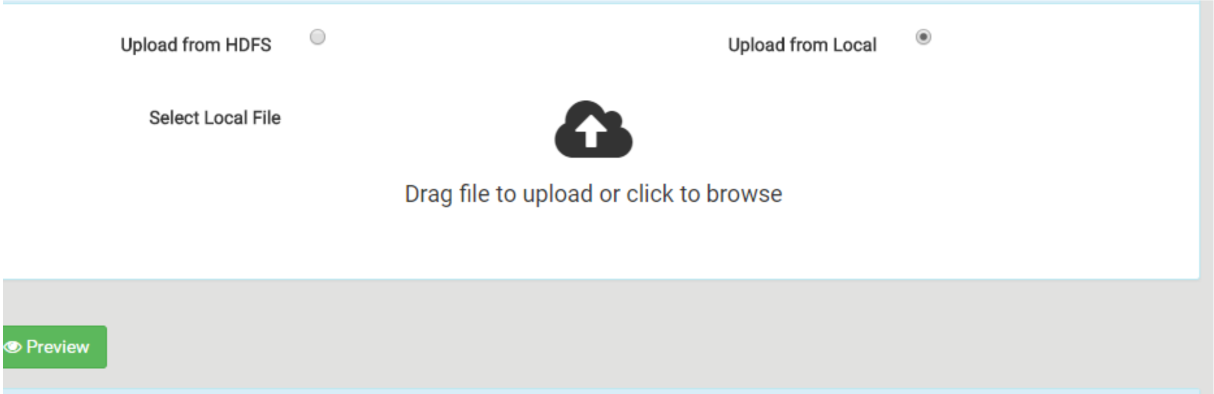


DATABASE
Select or search database/schema
fraud_detection

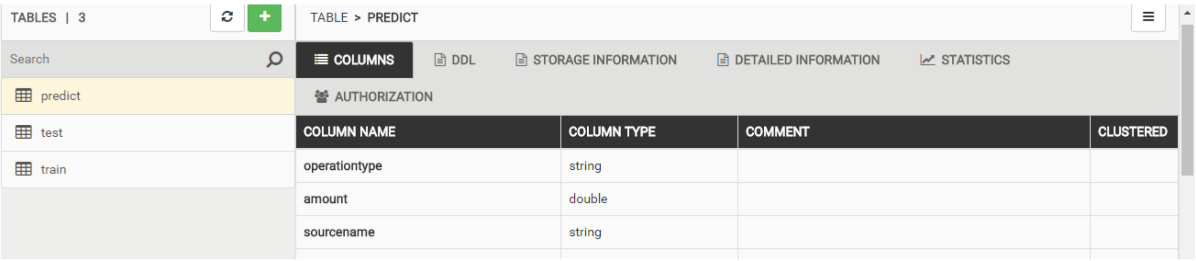
Name: Table name

COLUMN NAME	DATA TYPE	SIZE	ADVANCED	ACTION
+ Add New Column				

We drag our file .csv

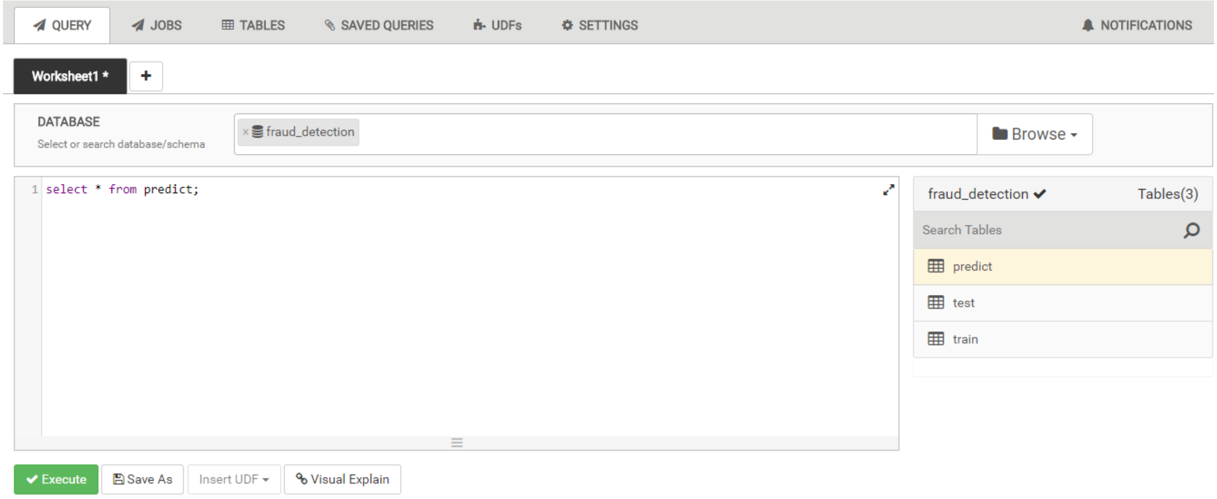


Here we find our tables.



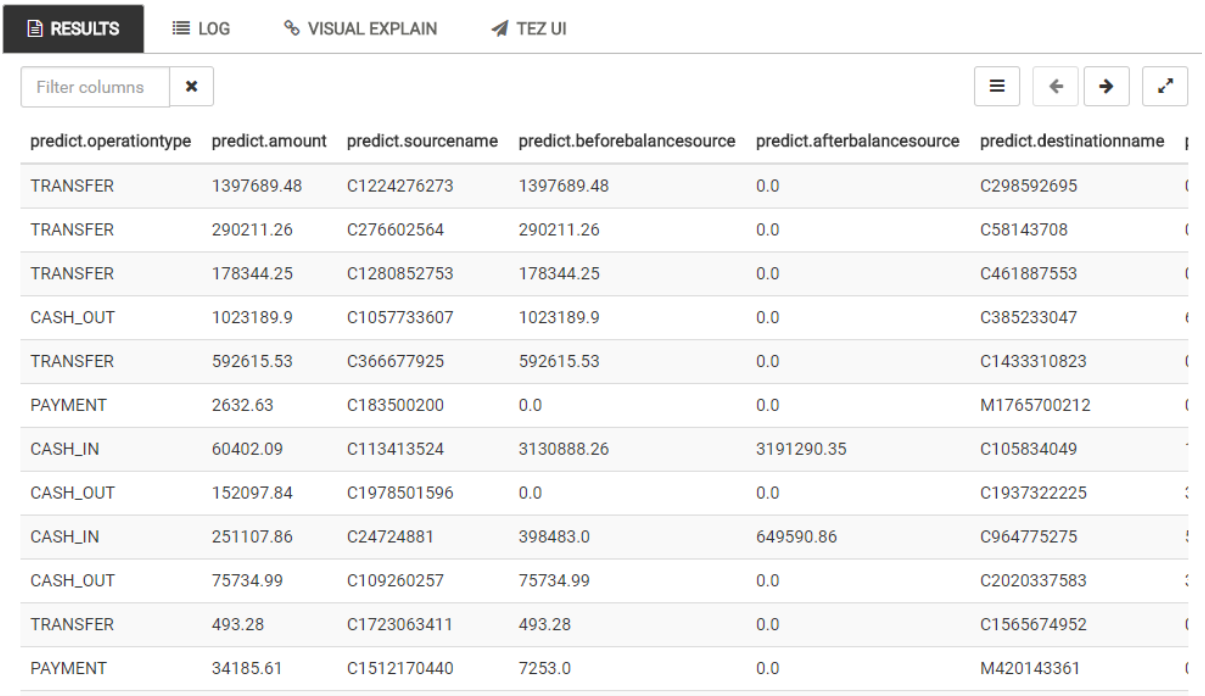
COLUMN NAME	COLUMN TYPE	COMMENT	CLUSTERED
operationtype	string		
amount	double		
sourcename	string		

Selection request :



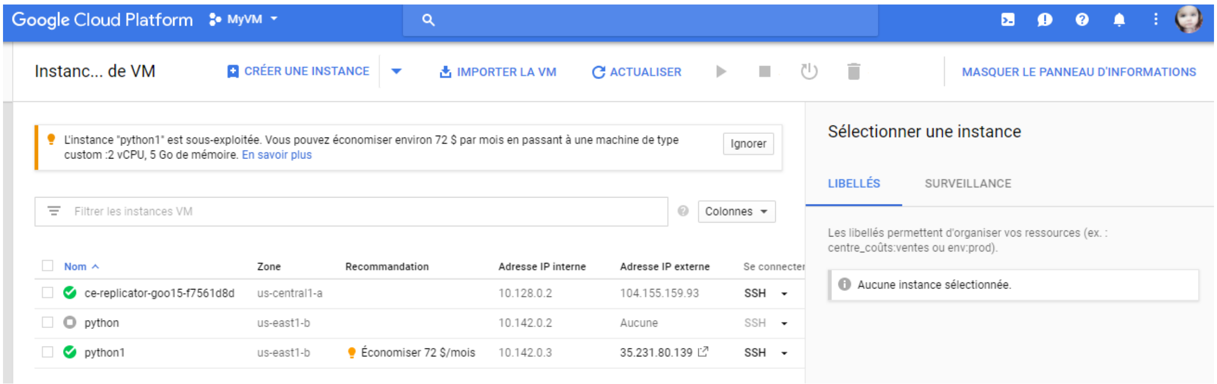
```
select * from predict;
```

Results:



predict.operationtype	predict.amount	predict.sourcename	predict.beforebalancesource	predict.afterbalancesource	predict.destinationname
TRANSFER	1397689.48	C1224276273	1397689.48	0.0	C298592695
TRANSFER	290211.26	C276602564	290211.26	0.0	C58143708
TRANSFER	178344.25	C1280852753	178344.25	0.0	C461887553
CASH_OUT	1023189.9	C1057733607	1023189.9	0.0	C385233047
TRANSFER	592615.53	C366677925	592615.53	0.0	C1433310823
PAYMENT	2632.63	C183500200	0.0	0.0	M1765700212
CASH_IN	60402.09	C113413524	3130888.26	3191290.35	C105834049
CASH_OUT	152097.84	C1978501596	0.0	0.0	C1937322225
CASH_IN	251107.86	C24724881	398483.0	649590.86	C964775275
CASH_OUT	75734.99	C109260257	75734.99	0.0	C2020337583
TRANSFER	493.28	C1723063411	493.28	0.0	C1565674952
PAYMENT	34185.61	C1512170440	7253.0	0.0	M420143361

We also create a VM in Google cloud to execute python script.

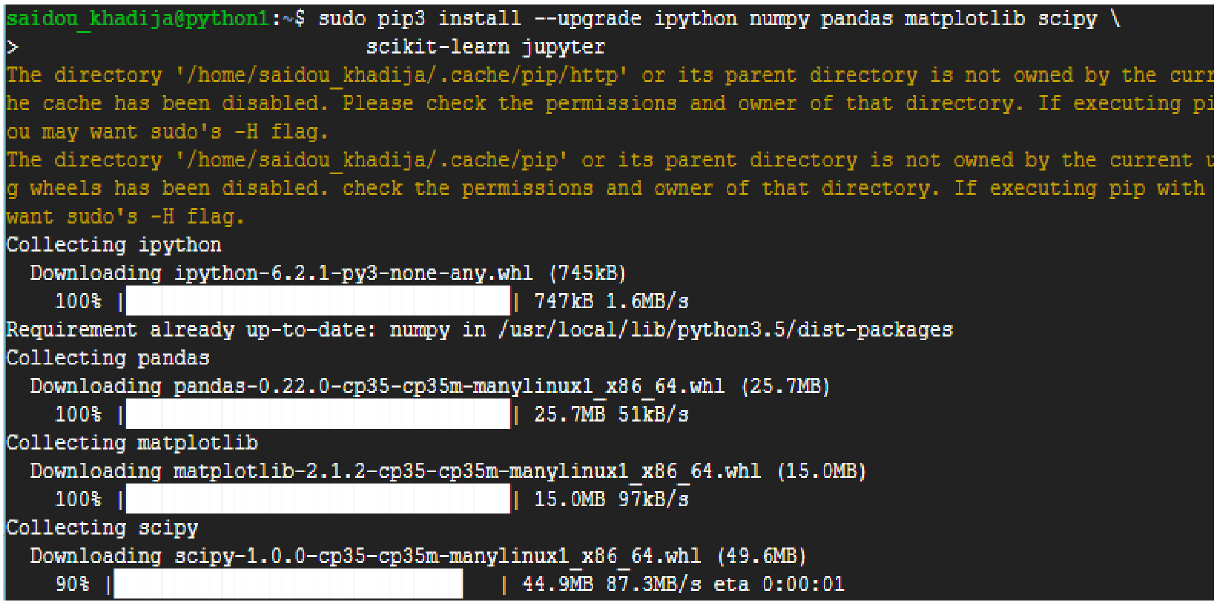


L'instance "python1" est sous-exploitée. Vous pouvez économiser environ 72 $ par mois en passant à une machine de type custom :2 vCPU, 5 Go de mémoire. En savoir plus

Nom	Zone	Recommandation	Adresse IP interne	Adresse IP externe	Se connecter
✓ ce-replicator-goo15-f7561d8d	us-central1-a		10.128.0.2	104.155.159.93	SSH
○ python	us-east1-b		10.142.0.2	Aucune	SSH
✓ python1	us-east1-b	Économiser 72 $/mois	10.142.0.3	35.231.80.139	SSH

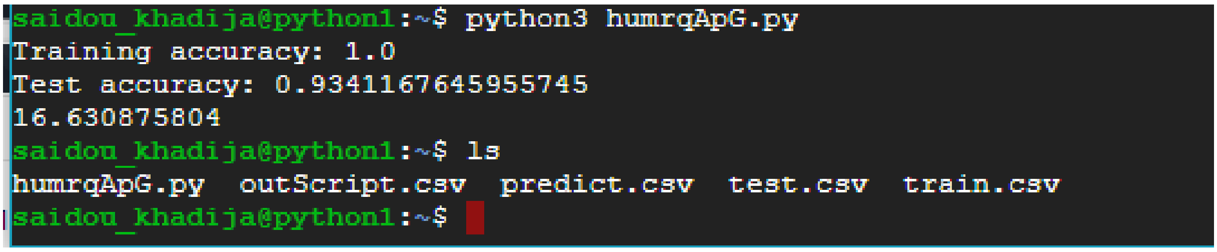
Les libellés permettent d'organiser vos ressources (ex. : centre_coûts:ventes ou env:prod).

Aucune instance sélectionnée.

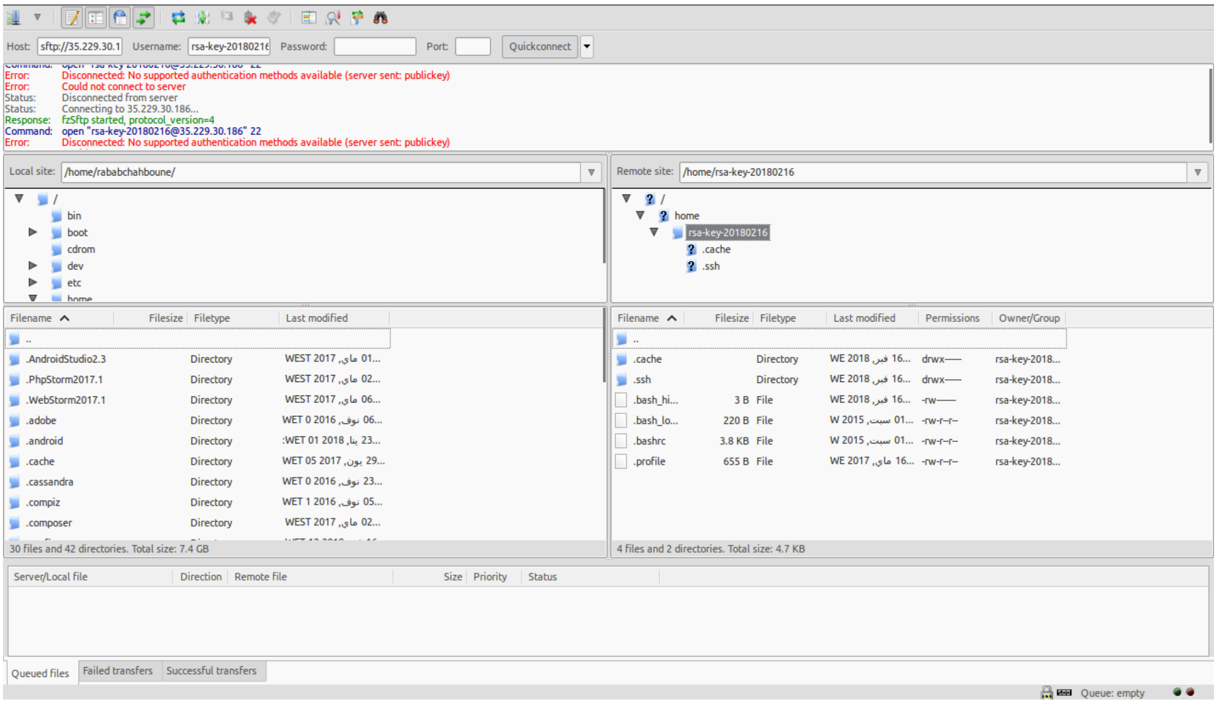


```
saidou_khadija@python1:~$ sudo pip3 install --upgrade ipython numpy pandas matplotlib scipy \
>                               scikit-learn jupyter
The directory '/home/saidou_khadija/.cache/pip/http' or its parent directory is not owned by the curr
he cache has been disabled. Please check the permissions and owner of that directory. If executing pi
ou may want sudo's -H flag.
The directory '/home/saidou_khadija/.cache/pip' or its parent directory is not owned by the current u
g wheels has been disabled. check the permissions and owner of that directory. If executing pip with
want sudo's -H flag.
Collecting ipython
  Downloading ipython-6.2.1-py3-none-any.whl (745kB)
    100% |████████████████████████████████| 747kB 1.6MB/s
Requirement already up-to-date: numpy in /usr/local/lib/python3.5/dist-packages
Collecting pandas
  Downloading pandas-0.22.0-cp35-cp35m-manylinux1_x86_64.whl (25.7MB)
    100% |████████████████████████████████| 25.7MB 51kB/s
Collecting matplotlib
  Downloading matplotlib-2.1.2-cp35-cp35m-manylinux1_x86_64.whl (15.0MB)
    100% |████████████████████████████████| 15.0MB 97kB/s
Collecting scipy
  Downloading scipy-1.0.0-cp35-cp35m-manylinux1_x86_64.whl (49.6MB)
    90% |█████████████████████████████   | 44.9MB 87.3MB/s eta 0:00:01
```

To send the predict.csv , test.csv, and train csv, we used FileZela:



```
saidou_khadija@python1:~$ python3 humrqApG.py
Training accuracy: 1.0
Test accuracy: 0.9341167645955745
16.630875804
saidou_khadija@python1:~$ ls
humrqApG.py  outScript.csv  predict.csv  test.csv  train.csv
saidou_khadija@python1:~$
```



Host: sftp://35.229.30.1 Username: rsa-key-2018021€ Password: Port:

```
Error:     Disconnected: No supported authentication methods available (server sent: publickey)
Error:     Could not connect to server
Status:    Disconnected from server
Status:    Connecting to 35.229.30.186...
Response:  fzSftp started, protocol_version=4
Command:   open "rsa-key-20180216@35.229.30.186" 22
Error:     Disconnected: No supported authentication methods available (server sent: publickey)
```

Local site: /home/rababchahboune/

Remote site: /home/rsa-key-20180216

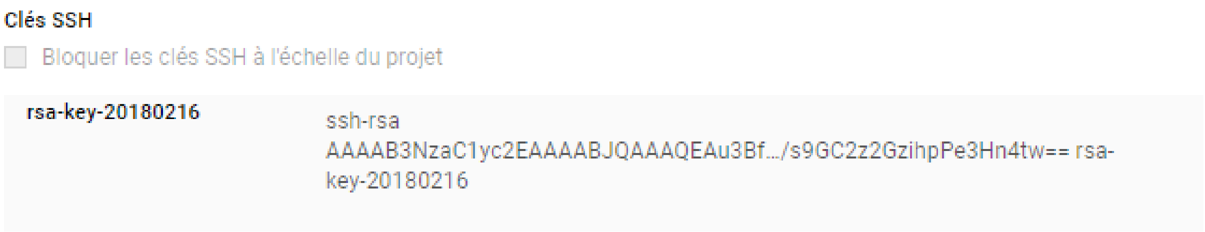
Filename	Filesize	Filetype	Last modified
..			
.AndroidStudio2.3		Directory	WEST 2017 ,ماي 01...
.PhpStorm2017.1		Directory	WEST 2017 ,ماي 02...
.WebStorm2017.1		Directory	WEST 2017 ,ماي 06...
.adobe		Directory	WET 0 2016 ,نوف 06...
.android		Directory	:WET 01 2018 ,ينا 23...
.cache		Directory	WET 05 2017 ,يون 29...
.cassandra		Directory	WET 0 2016 ,نوف 23...
.compiz		Directory	WET 1 2016 ,نوف 05...
.composer		Directory	WEST 2017 ,ماي 02...

30 files and 42 directories. Total size: 7.4 GB

Filename	Filesize	Filetype	Last modified	Permissions	Owner/Group
..					
.cache		Directory	WE 2018 ,فبر 16...	drwx------	rsa-key-2018...
.ssh		Directory	WE 2018 ,فبر 16...	drwx------	rsa-key-2018...
.bash_hi...	3 B	File	WE 2018 ,فبر 16...	-rw-------	rsa-key-2018...
.bash_lo...	220 B	File	W 2015 ,سبت 01...	-rw-r--r--	rsa-key-2018...
.bashrc	3.8 KB	File	W 2015 ,سبت 01...	-rw-r--r--	rsa-key-2018...
.profile	655 B	File	WE 2017 ,ماي 16...	-rw-r--r--	rsa-key-2018...

4 files and 2 directories. Total size: 4.7 KB

To connect to the machine with SSH we generated a key with puttygen then we copied this key to th ssh configuration machine:

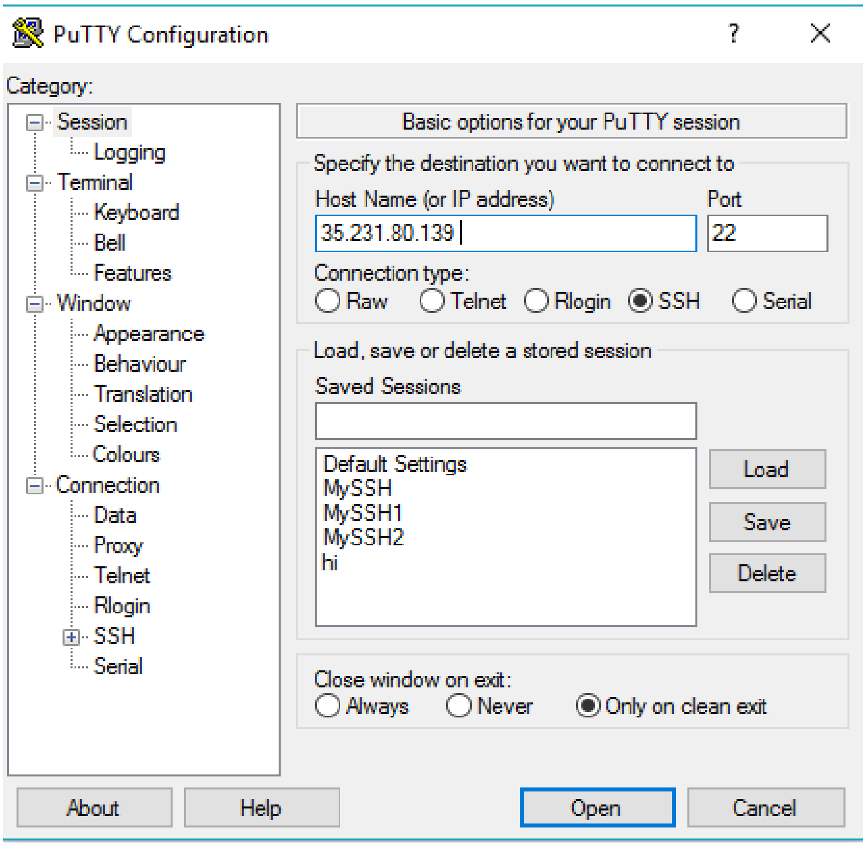


Clés SSH

□ Bloquer les clés SSH à l'échelle du projet

rsa-key-20180216	ssh-rsa AAAAB3NzaC1yc2EAAAABJQAAAQEAu3Bf.../s9GC2z2GzihpPe3Hn4tw== rsa-key-20180216

Then we importe the ppk in putty:



Specify the destination you want to connect to

Host Name (or IP address): 35.231.80.139

Port: 22

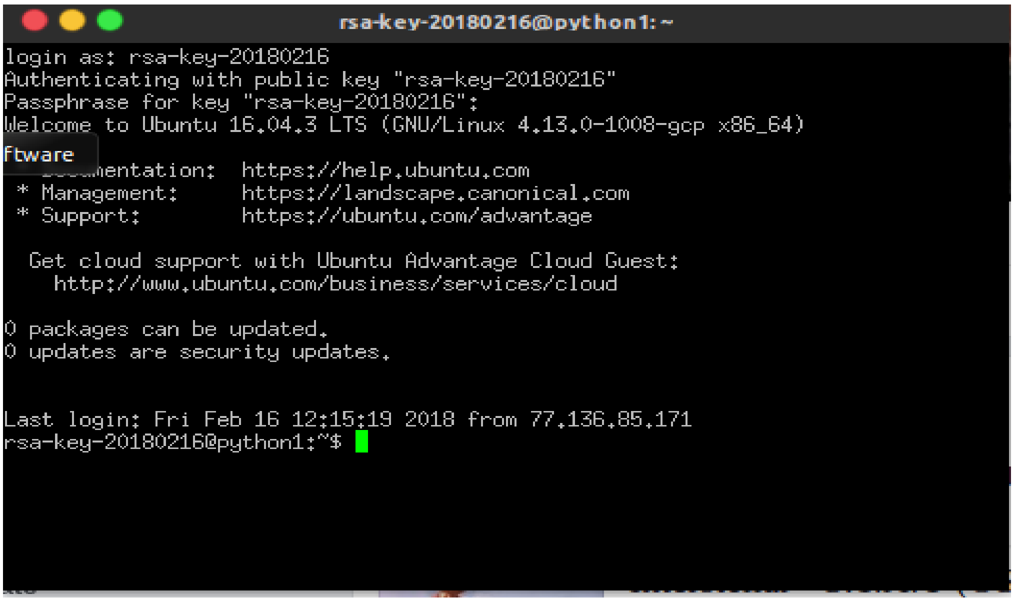
Connection type: ○ Raw ○ Telnet ○ Rlogin ● SSH ○ Serial

Load, save or delete a stored session

Saved Sessions

Default Settings
MySSH
MySSH1
MySSH2
hi

Close window on exit: ○ Always ○ Never ● Only on clean exit



```
login as: rsa-key-20180216
Authenticating with public key "rsa-key-20180216"
Passphrase for key "rsa-key-20180216":
Welcome to Ubuntu 16.04.3 LTS (GNU/Linux 4.13.0-1008-gcp x86_64)
ftware
   mentation:  https://help.ubuntu.com
 * Management:     https://landscape.canonical.com
 * Support:        https://ubuntu.com/advantage

  Get cloud support with Ubuntu Advantage Cloud Guest:
    http://www.ubuntu.com/business/services/cloud

0 packages can be updated.
0 updates are security updates.


Last login: Fri Feb 16 12:15:19 2018 from 77.136.85.171
rsa-key-20180216@python1:~$ 
```