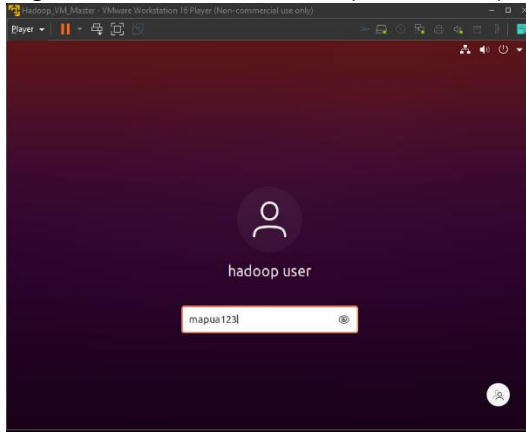


## Steps

1. Login to VM. Username: hduser, password: mapua123



2. Start hiveserver2

```
hduser@ubuntu:~$ hiveserver2
2021-02-10 17:14:19: Starting HiveServer2
Hive Session ID = 9861720f-5080-46d9-b0d7-1c7fcd0ddfaf
Hive Session ID = 5d1d0929-95c8-4a72-bbd5-082324b35a21
Hive Session ID = aa4379ad-b2e6-422e-9df4-c3e0a9833d0b
Hive Session ID = ada63f7b-3d88-46d4-a477-7054aaf7b885
Hive Session ID = 9eefde20-5c54-4969-b0bb-0666f7c917b8
Hive Session ID = f0f977f2-1bcc-474b-87cf-105811e57ec5
Hive Session ID = d883f7de-f706-48a9-80a0-a5e81d7f72c8
Hive Session ID = 58bb2a18-aa22-4aca-bb16-d6cd9c6a1153
```

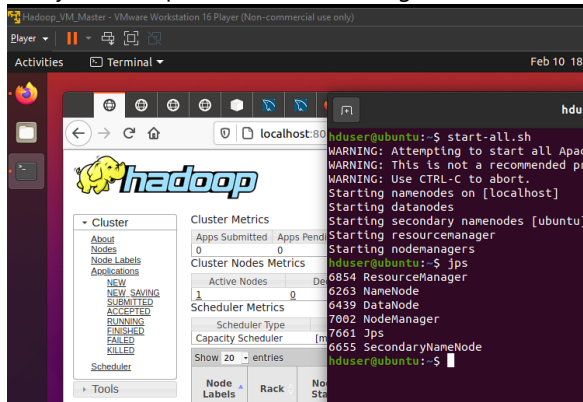
3. Check if mysql is running and run beeline

```
hduser@ubuntu:~$ ps -ef | grep mysql | grep -v grep
mysql      1099      1  0 17:10 ?        00:00:03
hduser@ubuntu:~$ beeline
Beeline version 3.1.2 by Apache Hive
beeline> select * from customer;
No current connection
beeline>
```

4. Run Hadoop daemons

```
hduser@ubuntu:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as h
WARNING: This is not a recommended production deployment co
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [ubuntu]
Starting resourcemanager
Starting nodemanagers
hduser@ubuntu:~$
```

5. Verify if Hadoop daemons are running

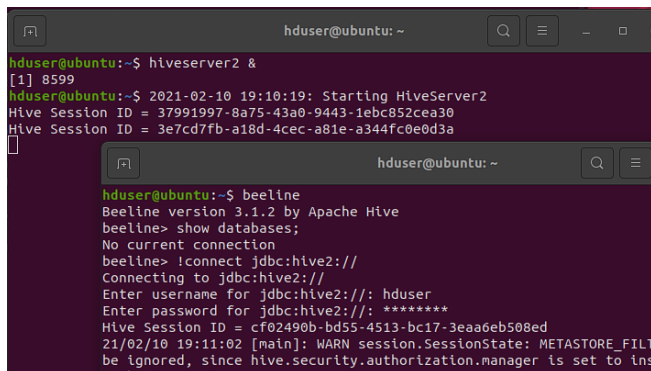


6. Open another terminal and run hiveserver and beeline. Connect to hive and provide the below username and password.

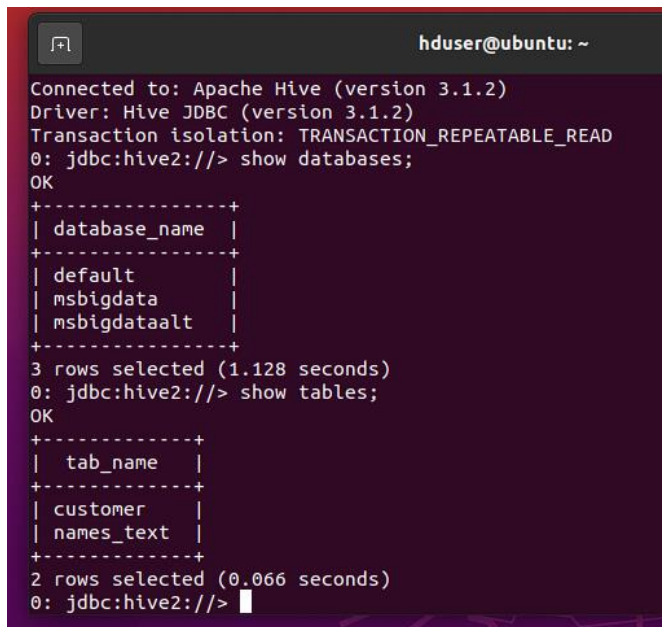
command: !connect jdbc:hive2://

username: hduser

password: mapua123



7. Run the commands show databases and show tables as shown.



8. Verify if you can see the customer table in the Hadoop file system.

```
hduser@ubuntu: ~  
hduser@ubuntu:~$ hadoop fs -ls /user/hive/warehouse  
Found 8 items  
drwxr-xr-x - hduser supergroup 0 2021-01-24 04:56 /user/hive/warehouse/SHOPPING.db  
-rw-r--r-- 1 hduser supergroup 133 2021-02-10 03:24 /user/hive/warehouse/cust.txt  
drwxr-xr-x - hduser supergroup 0 2021-02-09 08:15 /user/hive/warehouse/customer  
drwxr-xr-x - hduser supergroup 0 2021-01-30 02:57 /user/hive/warehouse/employee  
drwxr-xr-x - hduser supergroup 0 2021-02-09 04:00 /user/hive/warehouse/msbigdata.db  
drwxr-xr-x - hduser supergroup 0 2021-01-24 00:27 /user/hive/warehouse/mydatabase.db  
drwxr-xr-x - hduser supergroup 0 2021-01-27 00:12 /user/hive/warehouse/newdb.db  
drwxr-xr-x - hduser supergroup 0 2021-01-23 23:36 /user/hive/warehouse/std_db.db  
hduser@ubuntu:~$
```

9. Run the HiveQL select \* from customer. You should see the below display

```
0: jdbc:hive2://> select * from customer;  
OK  
+-----+-----+-----+-----+  
| customer.cust_id | customer.first_name | customer.last_name | customer.email_address |  
+-----+-----+-----+-----+  
| 34567678 | Mary | Jones | mary.jones@isp.com |  
| 897572388 | Harry | Schmidt | harry.schmidt@isp.com |  
| 89976576 | Tom | Smith | thomas.smith@another_isp.com |  
+-----+-----+-----+-----+  
3 rows selected (1.909 seconds)  
0: jdbc:hive2://>
```

10. Create and activate softDes virtual environment with the below commands.

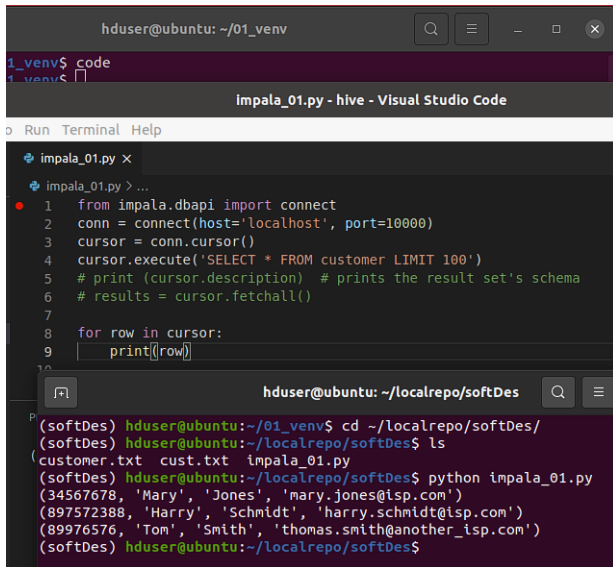
- virtualenv softDes
- source softDes/bin/activate

```
hduser@ubuntu: ~/01_venv  
hduser@ubuntu:~/01_venv$ ls  
softDes tf  
hduser@ubuntu:~/01_venv$ source softDes/bin/activate  
(softDes) hduser@ubuntu:~/01_venv$
```

11. install impyla in the current environment

```
hduser@ubuntu: ~/01_venv  
hduser@ubuntu:~/01_venv$ ls  
softDes tf  
hduser@ubuntu:~/01_venv$ source softDes/bin/activate  
(softDes) hduser@ubuntu:~/01_venv$ pip install impyla  
Collecting impyla  
Using cached impyla-0.16.3-py3-none-any.whl  
Collecting bitarray  
Using cached bitarray-1.6.3-cp38-cp38-linux_x86_64.whl  
Collecting six  
Using cached six-1.15.0-py2.py3-none-any.whl (10 kB)  
Collecting thriftpy2<0.5.0,>=0.4.0  
Using cached thriftpy2-0.4.14-cp38-cp38-linux_x86_64.whl  
Collecting ply<4.0,>=3.4  
Using cached ply-3.11-py2.py3-none-any.whl (49 kB)  
Installing collected packages: ply, thriftpy2, six, bitarray, impyla  
Successfully installed bitarray-1.6.3 impyla-0.16.3 ply-3.11 six-1.15.0  
(softDes) hduser@ubuntu:~/01_venv$
```

12. Run Visual Studio Code and type the following Python program. You should see the below update.



```
hduser@ubuntu: ~/01_venv
1_venv$ code
1_venv$

impala_01.py - hive - Visual Studio Code

Run Terminal Help

impala_01.py X
impala_01.py > ...
1 from impala.dbapi import connect
2 conn = connect(host='localhost', port=10000)
3 cursor = conn.cursor()
4 cursor.execute('SELECT * FROM customer LIMIT 100')
5 # print (cursor.description) # prints the result set's schema
6 # results = cursor.fetchall()
7
8 for row in cursor:
9     print(row)
10
11
hduser@ubuntu: ~/localrepo/softDes
P
(softDes) hduser@ubuntu:~/01_venv$ cd ~/localrepo/softDes/
(softDes) hduser@ubuntu:~/localrepo/softDes$ ls
customer.txt cust.txt impala_01.py
(softDes) hduser@ubuntu:~/localrepo/softDes$ python impala_01.py
(34567678, 'Mary', 'Jones', 'mary.jones@isp.com')
(897572388, 'Harry', 'Schmidt', 'harry.schmidt@isp.com')
(89976576, 'Tom', 'Smith', 'thomas.smith@another_isp.com')
(softDes) hduser@ubuntu:~/localrepo/softDes$
```

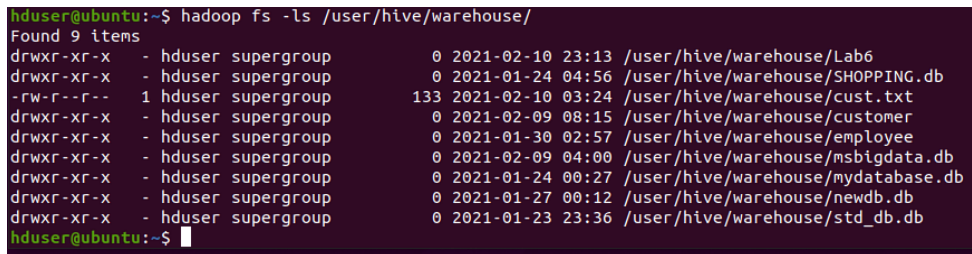
```
from impala.dbapi import connect
conn = connect(host='localhost', port=10000)
cursor = conn.cursor()
cursor.execute('SELECT * FROM customer LIMIT 100')
# print (cursor.description) # prints the result set's schema
# results = cursor.fetchall()
```

```
for row in cursor:
    print(row)
```

Try to run stop-all.sh in your terminal and verify if you can still see the p

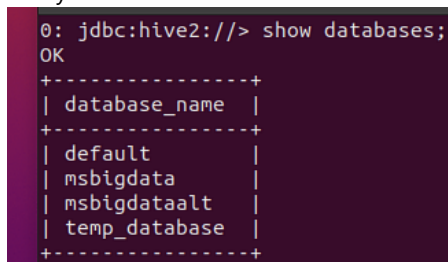
13. Create LR6 directory/folder in the Hadoop file system using the below command and verify if it was created successfully.

```
hadoop fs -mkdir /user/hive/warehouse/Lab6
```



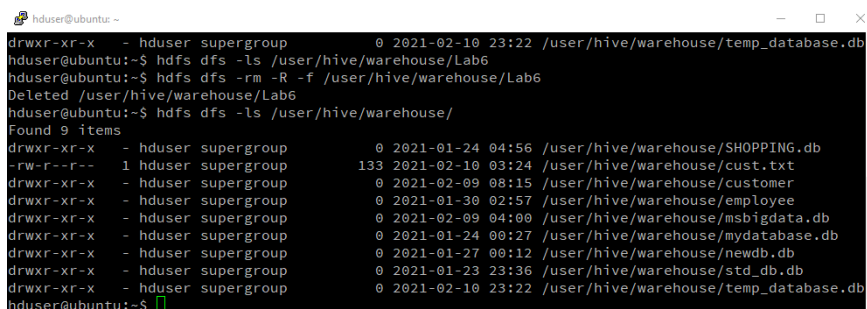
```
hduser@ubuntu:~$ hadoop fs -ls /user/hive/warehouse/
Found 9 items
drwxr-xr-x - hduser supergroup 0 2021-02-10 23:13 /user/hive/warehouse/Lab6
drwxr-xr-x - hduser supergroup 0 2021-01-24 04:56 /user/hive/warehouse/SHOPPING.db
-rw-r--r-- 1 hduser supergroup 133 2021-02-10 03:24 /user/hive/warehouse/cust.txt
drwxr-xr-x - hduser supergroup 0 2021-02-09 08:15 /user/hive/warehouse/customer
drwxr-xr-x - hduser supergroup 0 2021-01-30 02:57 /user/hive/warehouse/employee
drwxr-xr-x - hduser supergroup 0 2021-02-09 04:00 /user/hive/warehouse/msbigdata.db
drwxr-xr-x - hduser supergroup 0 2021-01-24 00:27 /user/hive/warehouse/mydatabase.db
drwxr-xr-x - hduser supergroup 0 2021-01-27 00:12 /user/hive/warehouse/newdb.db
drwxr-xr-x - hduser supergroup 0 2021-01-23 23:36 /user/hive/warehouse/std_db.db
hduser@ubuntu:~$
```

14. Create a new database temp\_database in the /user/hive/warehouse directory  
CREATE DATABASE temp\_database;  
15. Verify the database creation as shown.



```
0: jdbc:hive2://> show databases;
OK
+-----+
| database_name |
+-----+
| default       |
| msbigdata     |
| msbigdataalt  |
| temp_database |
+-----+
```

Verify using the -ls hdfs command.



```
hduser@ubuntu:~$ hdfs dfs -ls /user/hive/warehouse/temp_database.db
hduser@ubuntu:~$ hdfs dfs -ls /user/hive/warehouse/
Found 9 items
drwxr-xr-x - hduser supergroup 0 2021-01-24 04:56 /user/hive/warehouse/SHOPPING.db
-rw-r--r-- 1 hduser supergroup 133 2021-02-10 03:24 /user/hive/warehouse/cust.txt
drwxr-xr-x - hduser supergroup 0 2021-02-09 08:15 /user/hive/warehouse/customer
drwxr-xr-x - hduser supergroup 0 2021-01-30 02:57 /user/hive/warehouse/employee
drwxr-xr-x - hduser supergroup 0 2021-02-09 04:00 /user/hive/warehouse/msbigdata.db
drwxr-xr-x - hduser supergroup 0 2021-01-24 00:27 /user/hive/warehouse/mydatabase.db
drwxr-xr-x - hduser supergroup 0 2021-01-27 00:12 /user/hive/warehouse/newdb.db
drwxr-xr-x - hduser supergroup 0 2021-01-23 23:36 /user/hive/warehouse/std_db.db
drwxr-xr-x - hduser supergroup 0 2021-02-10 23:22 /user/hive/warehouse/temp_database.db
hduser@ubuntu:~$
```

Note: You can clear the beeline terminal by using the key combination Ctrl + L.

16. Run the below commands:

```
use temp_database;
```

```
create table temp_table (z int);
```

```
show tables;
```

You should have the same below display

```
0: jdbc:hive2://> show tables;
OK
+-----+
| tab_name |
+-----+
| temp_table |
+-----+
1 row selected (0.068 seconds)
0: jdbc:hive2://> Hive Session ID = 01934dd1-16c1-4df8-a
Hive Session ID = fa1edbd8-599b-493b-a7eb-e374fe0e2613
Hive Session ID = 65260859-9989-4a01-9cad-d8ea6d6b6d09
Hive Session ID = 0cad1220-09aa-487a-8698-546cace722ac
```

17. Ssh login to Master VM

```
hduser@ubuntu: ~
login as: hduser
hduser@192.168.92.152's password:
Welcome to Ubuntu 20.04.2 LTS (GNU/Linux 5.8.0-43-generic)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

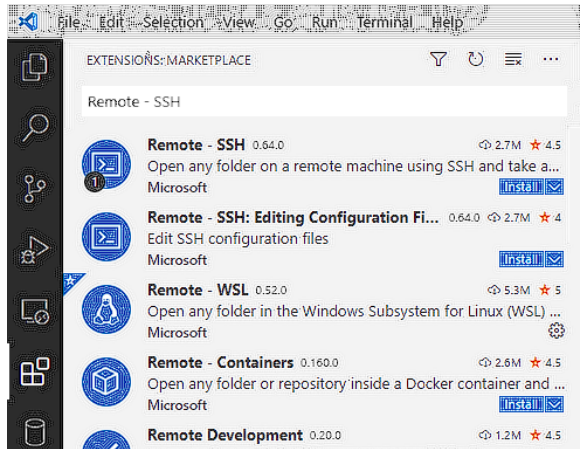
1 update can be installed immediately.
0 of these updates are security updates.
To see these additional updates run: apt list --upgradable

Your Hardware Enablement Stack (HWE) is supported until 2025-04-01
Last login: Fri Feb 12 22:19:27 2021 from 192.168.92.1
hduser@ubuntu:~$
```

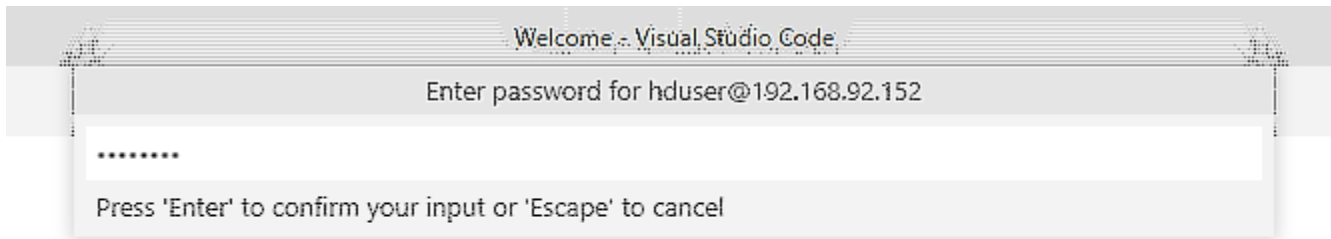
18.

```
hduser@ubuntu:~$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hduser/.ssh/id_rsa):
/home/hduser/.ssh/id_rsa already exists.
Overwrite (y/n)? y
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/hduser/.ssh/id_rsa
Your public key has been saved in /home/hduser/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:mNlpmgzjEsJINNIpi+LMIDKeS7eX3qgzcGxAsrx3ASY hduser@ubuntu
The key's randomart image is:
+---[RSA 3072]-----+
|. + . |
|XE+o |
|XB+ . |
|OB . . = . |
|B.B + * S |
| B O B + |
| * = * |
| + o . |
| .+ |
+---[SHA256]-----+
hduser@ubuntu:~$
```

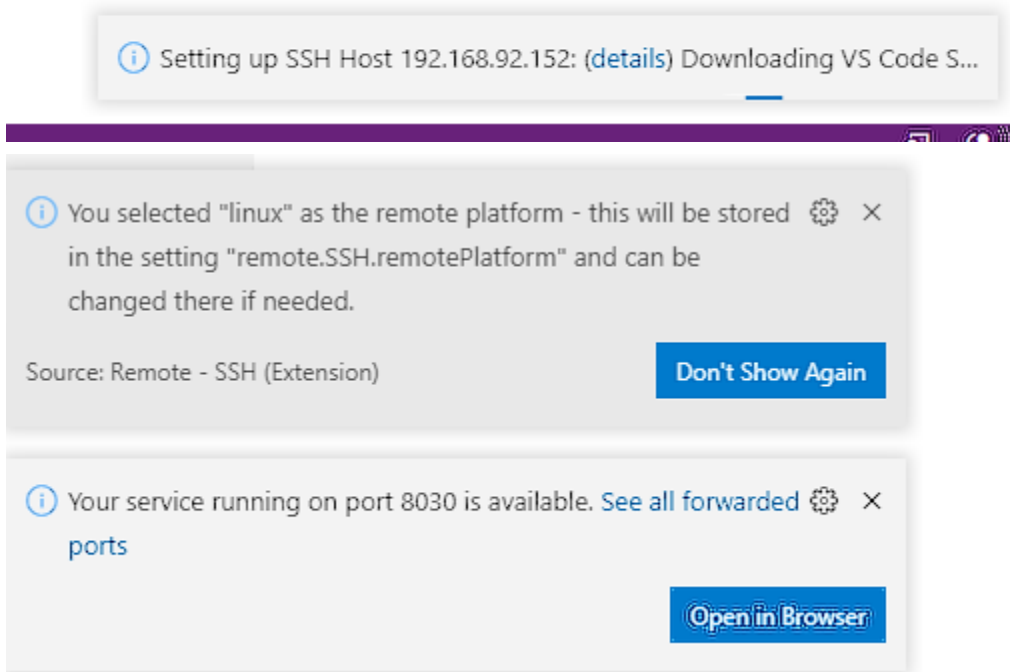
## 19. Install Remote-SSH on VS Code Host OS

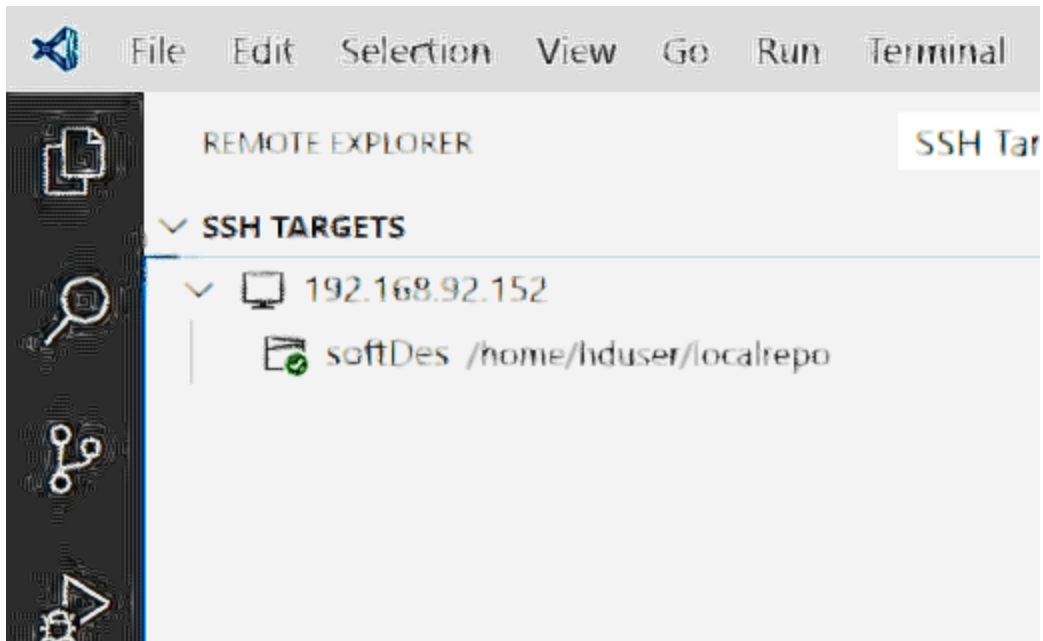


## 20. Sdfdf



node





## 21. Generate PuTTY key

