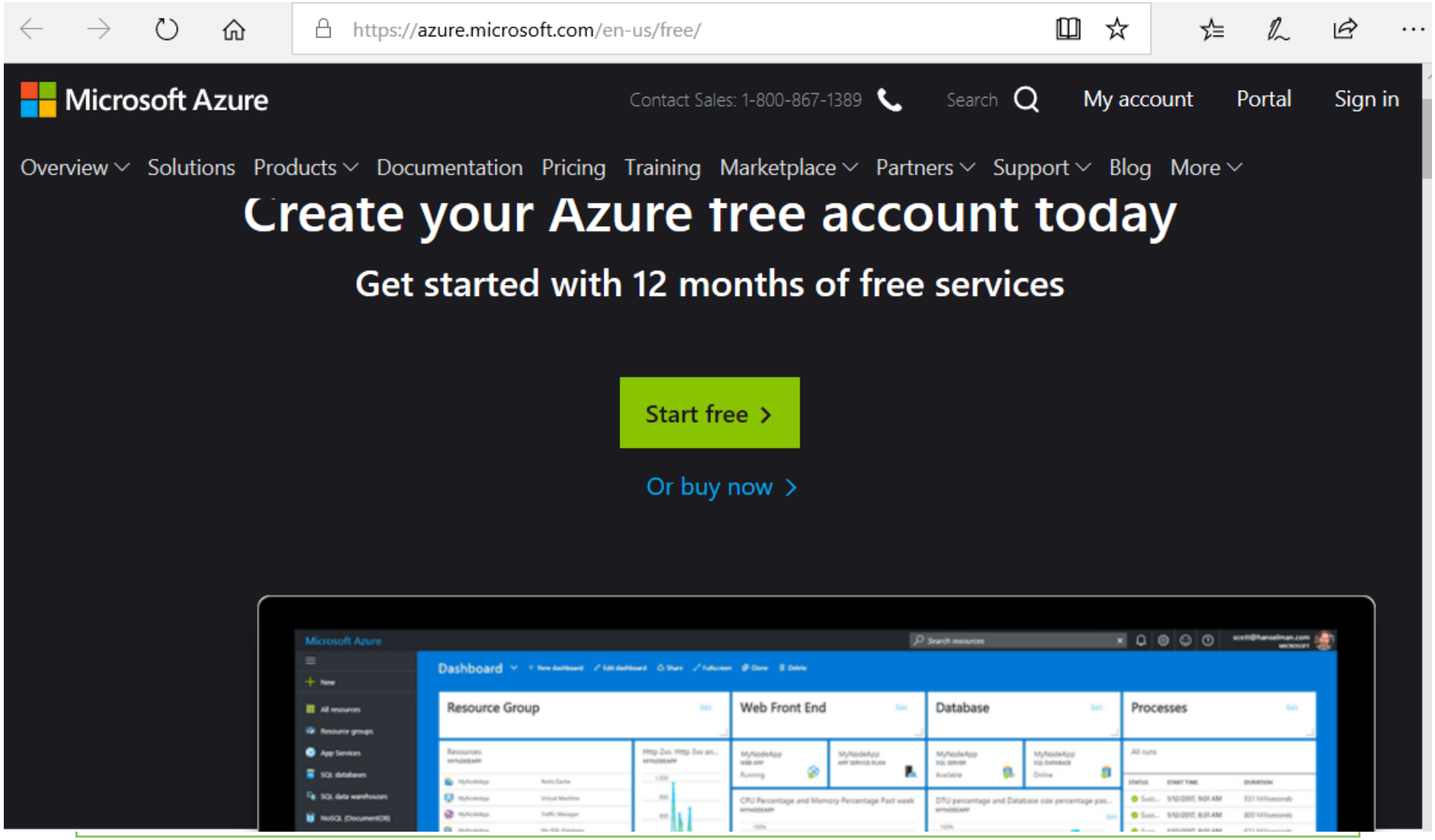
NYU | TANDON SCHOOL OF ENGINEERING

Microsoft Azure Cloud & Hortonworks Hadoop

Microsoft Azure

- Microsoft Azure is a cloud computing platform and infrastructure, created by Microsoft, for building, deploying and managing applications and services through a global network of Microsoft-managed and Microsoft partner hosted datacenters. It provides both PaaS and IaaS services and supports many different programming languages, tools and frameworks, including both Microsoft-specific and third-party software and systems. Azure was announced in October 2008 and released on 1 February 2010 as Windows Azure, before being renamed to Microsoft Azure on 25 March 2014
- **Create your trial account through the following link ($200 credit for one month):**
- **https://azure.microsoft.com/en-us/free/**

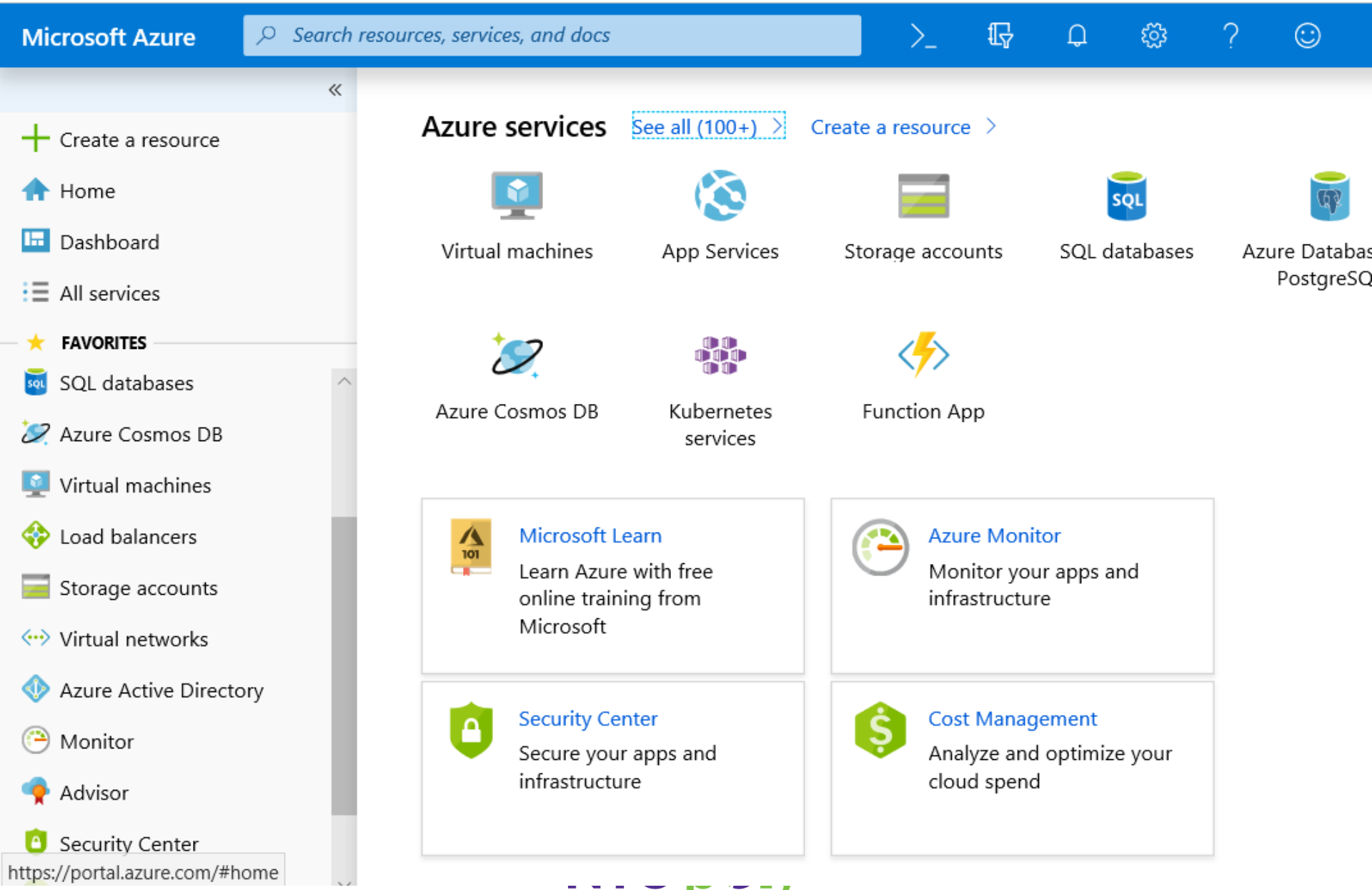


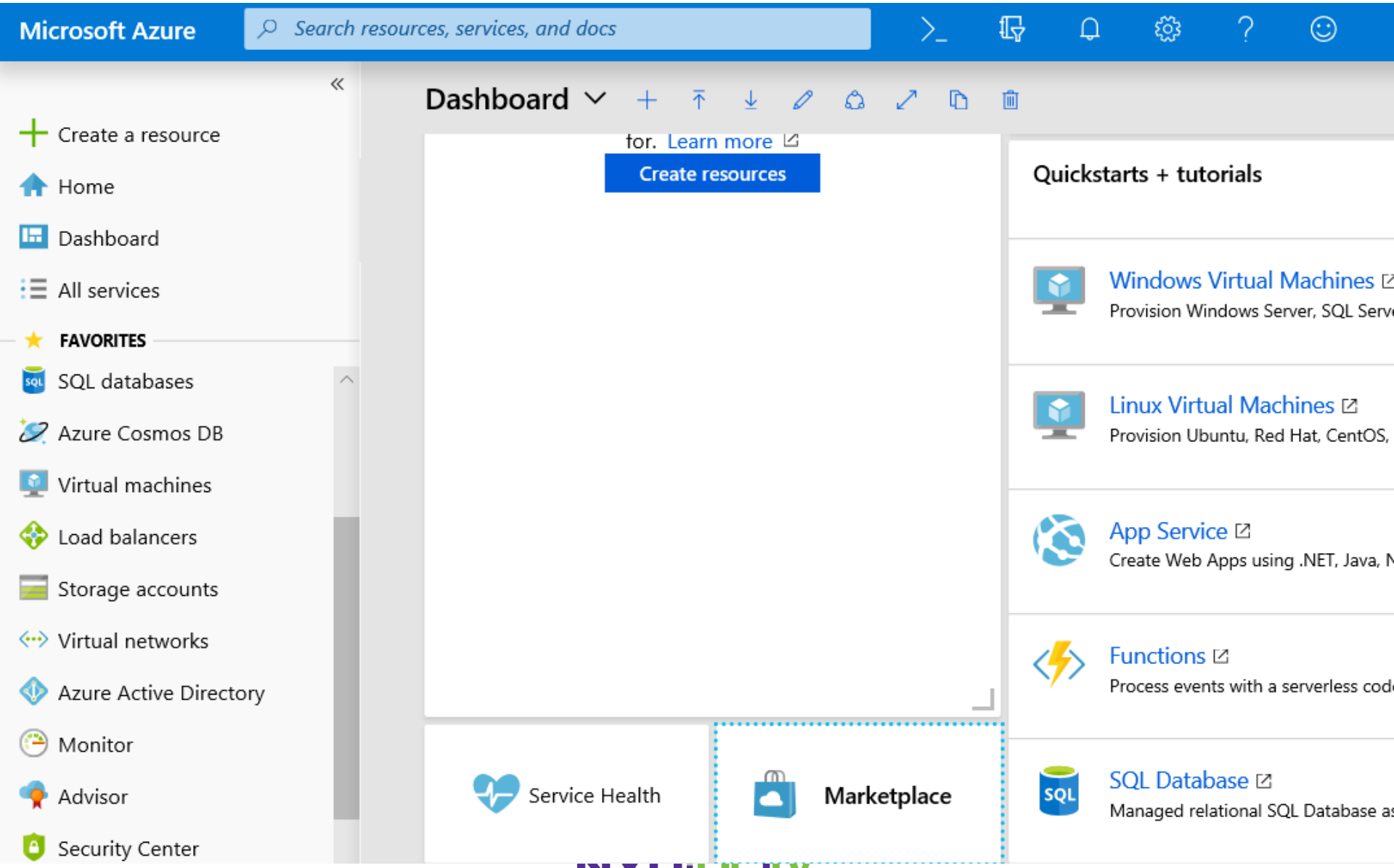
NEW YORK UNIVERSITY

4/30/19

NYU·poly
POLYTECHNIC INSTITUTE OF NEW YORK UNIVERSITY

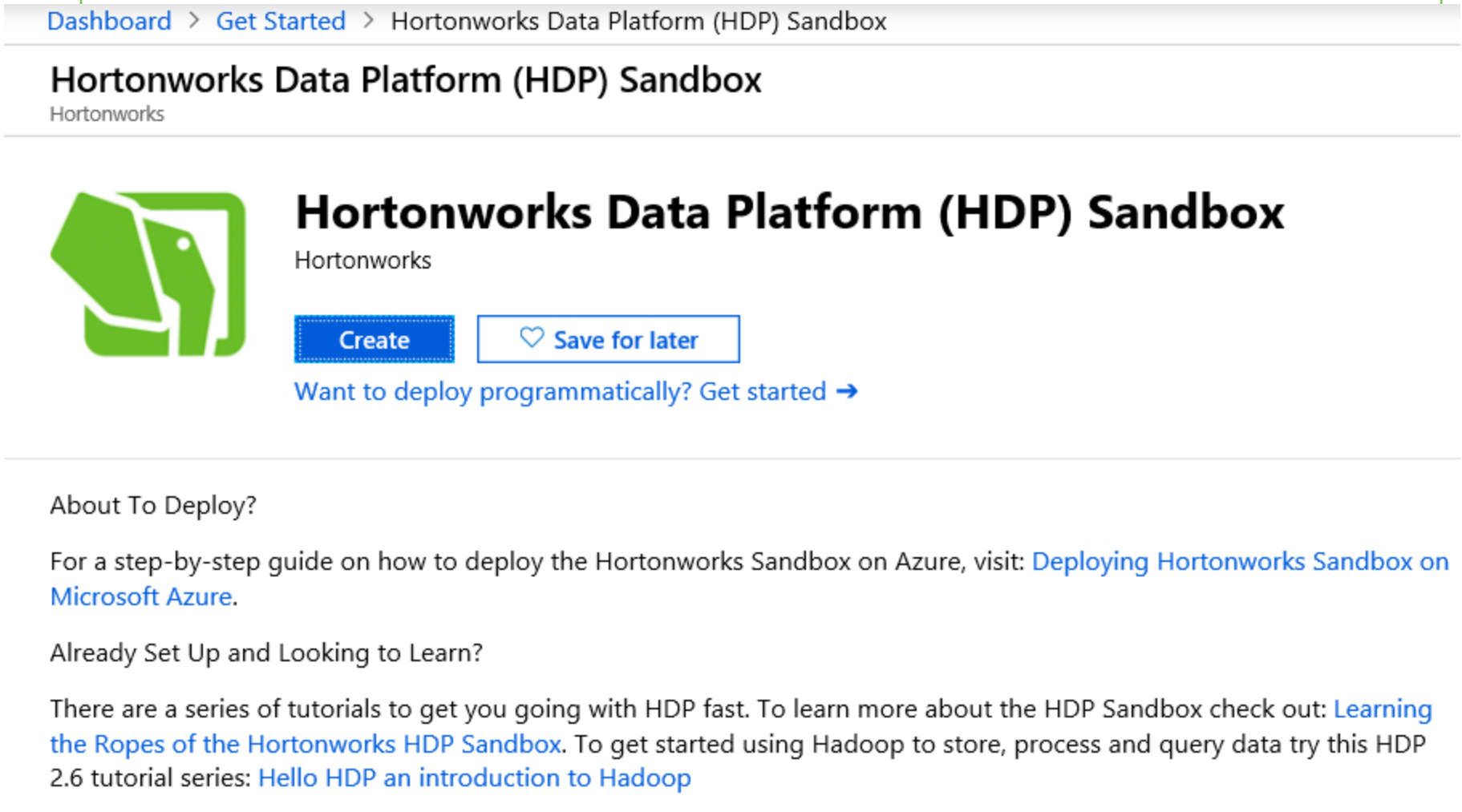
Leading invention, innovation and entrepreneurship


Microsoft Azure
Search resources, services, and docs
Create a resource
Home
Dashboard
All services
FAVORITES
SQL databases
Azure Cosmos DB
Virtual machines
Load balancers
Storage accounts
Virtual networks
Azure Active Directory
Monitor
Advisor
Security Center
https://portal.azure.com/#home
Azure services
See all (100+)
Create a resource
Virtual machines
App Services
Storage accounts
SQL databases
Azure Databas
PostgreSQ
Azure Cosmos DB
Kubernetes services
Function App
Microsoft Learn
Learn Azure with free online training from Microsoft
Azure Monitor
Monitor your apps and infrastructure
Security Center
Secure your apps and infrastructure
Cost Management
Analyze and optimize your cloud spend

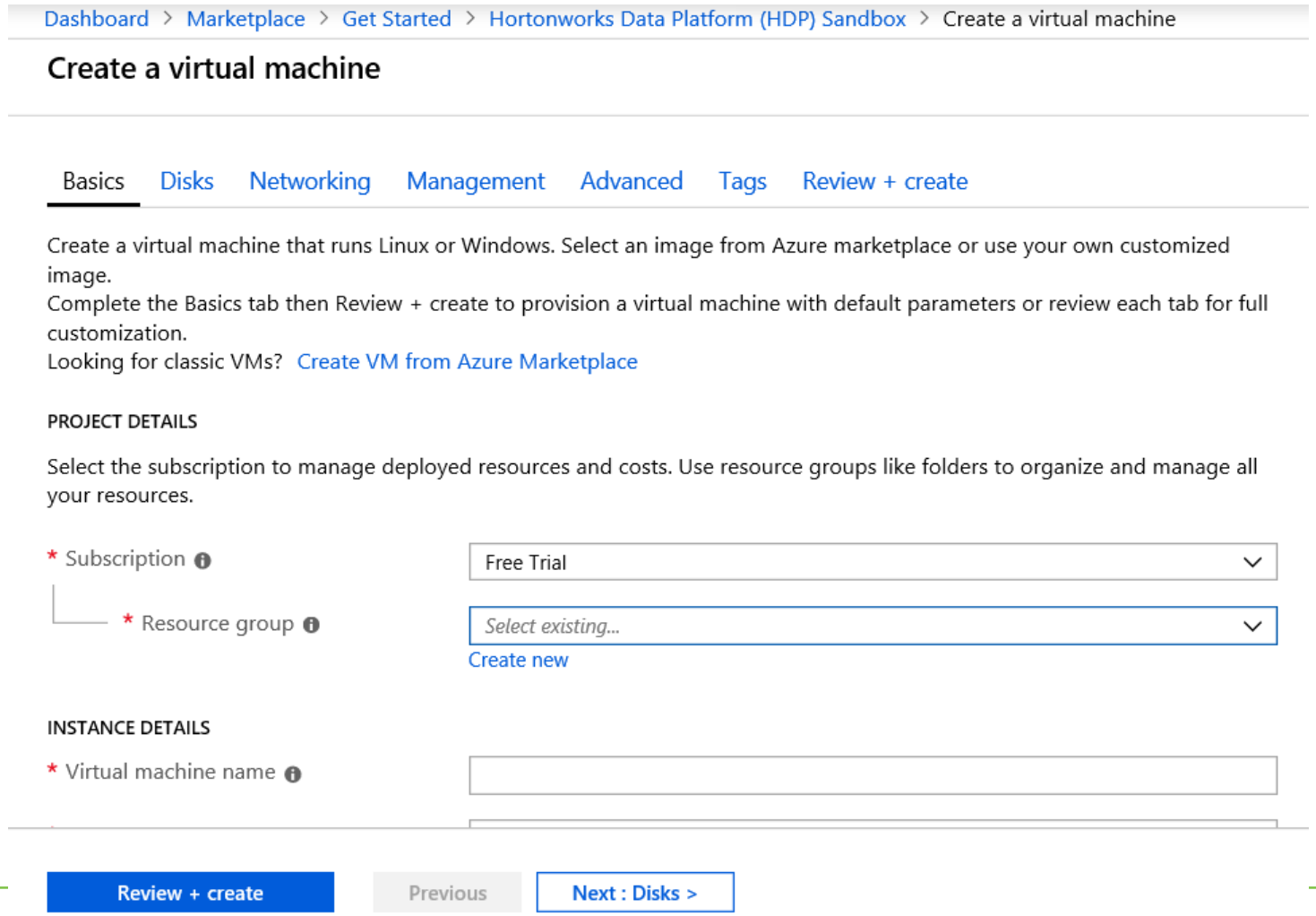

Microsoft Azure
Search resources, services, and docs
Dashboard
Create a resource
Home
Dashboard
All services
FAVORITES
SQL databases
Azure Cosmos DB
Virtual machines
Load balancers
Storage accounts
Virtual networks
Azure Active Directory
Monitor
Advisor
Security Center
for. Learn more
Create resources
Quickstarts + tutorials
Windows Virtual Machines
Provision Windows Server, SQL Serve
Linux Virtual Machines
Provision Ubuntu, Red Hat, CentOS,
App Service
Create Web Apps using .NET, Java, N
Functions
Process events with a serverless code
SQL Database
Managed relational SQL Database as
Service Health
Marketplace

Navigate to the 'MarketPlace'

Search for Hortonworks. Click on the Hortonworks Sandbox (HDP) Sandbox.



Click Create to Create a Virtual Machine

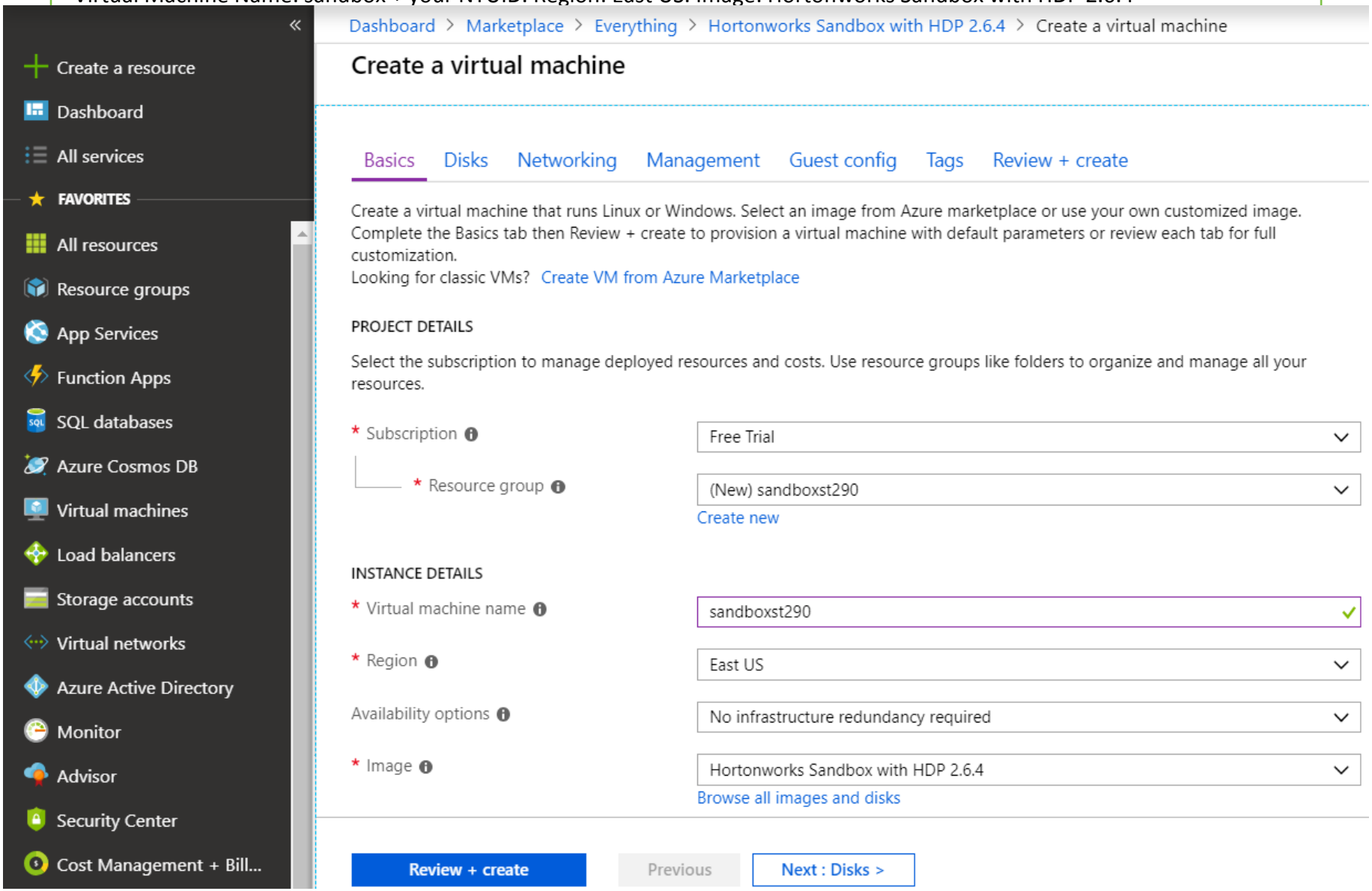


NYU·poly
POLYTECHNIC INSTITUTE OF NEW YORK UNIVERSITY

Create VM (Virtual Machine) for Hornworks Sandbox

Subscription: Free Trial. Resource Group: Create new, same as virtual machine name

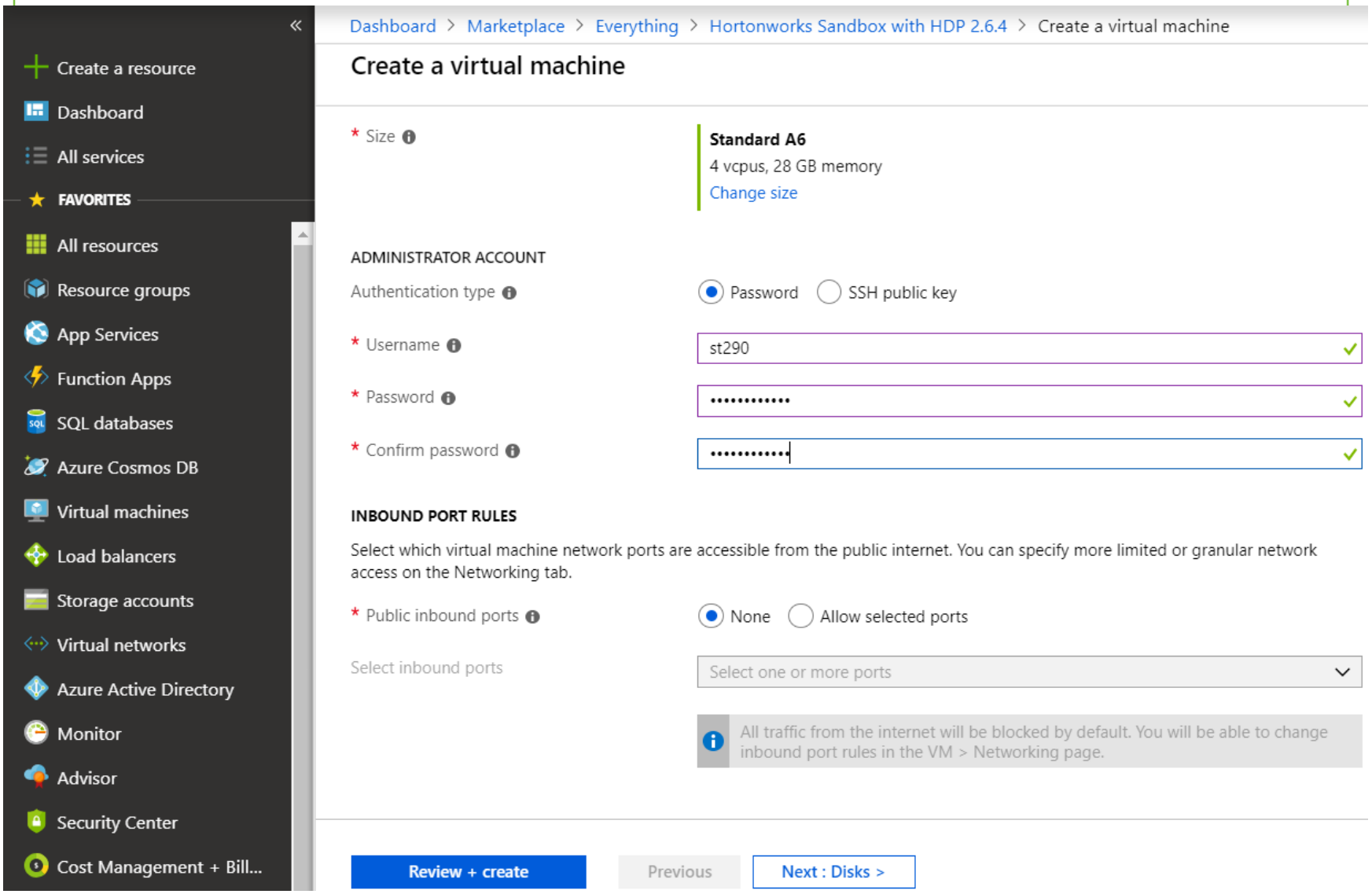
Virtual Machine Name: sandbox + your NYUID. Region: East US. Image: Hortonworks Sandbox with HDP 2.6.4



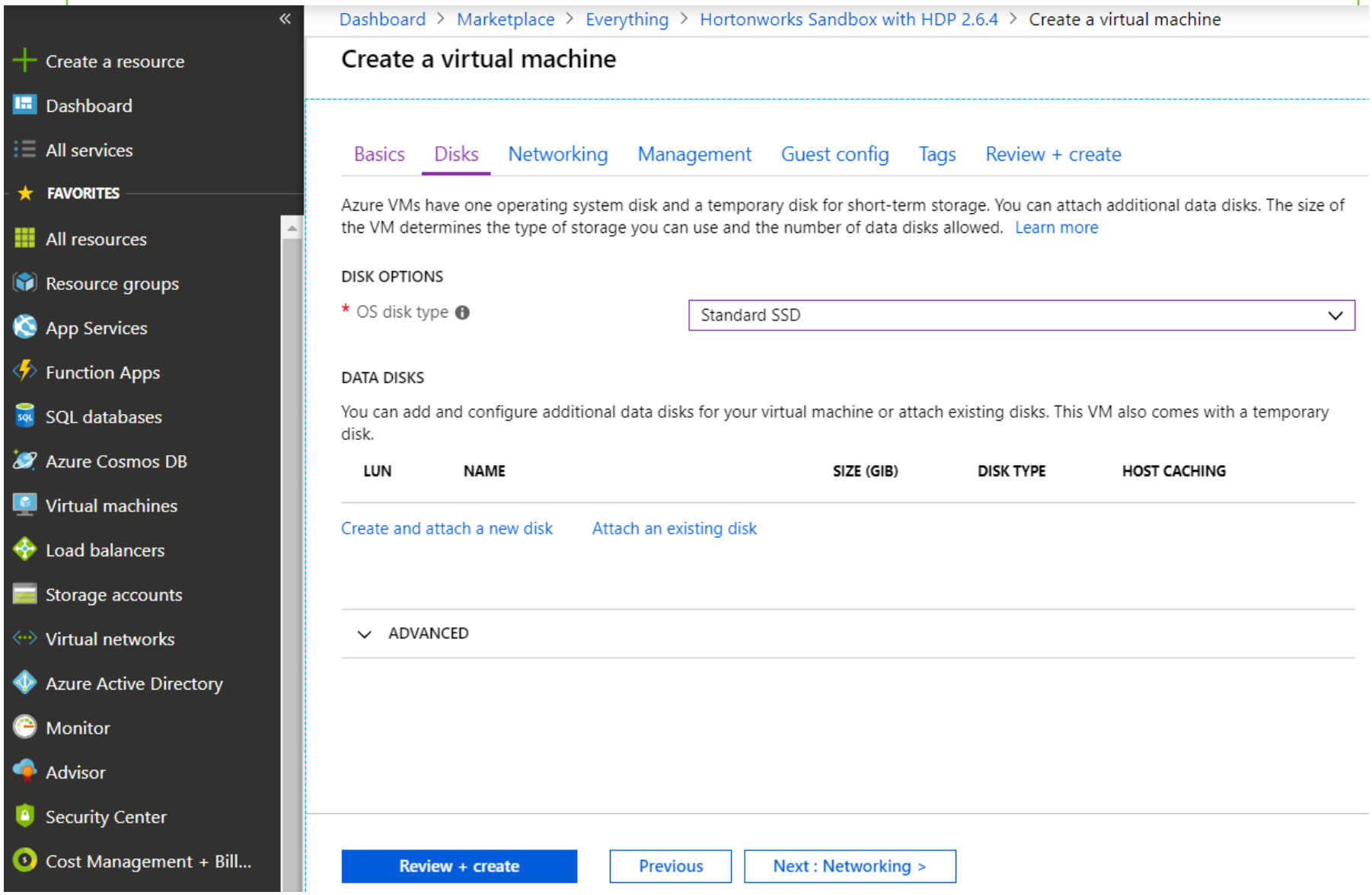
Create VM (Virtual Machine) for Hornworks Sandbox (Continue)

Select VM Size: Standard A6. Username: your NYUID

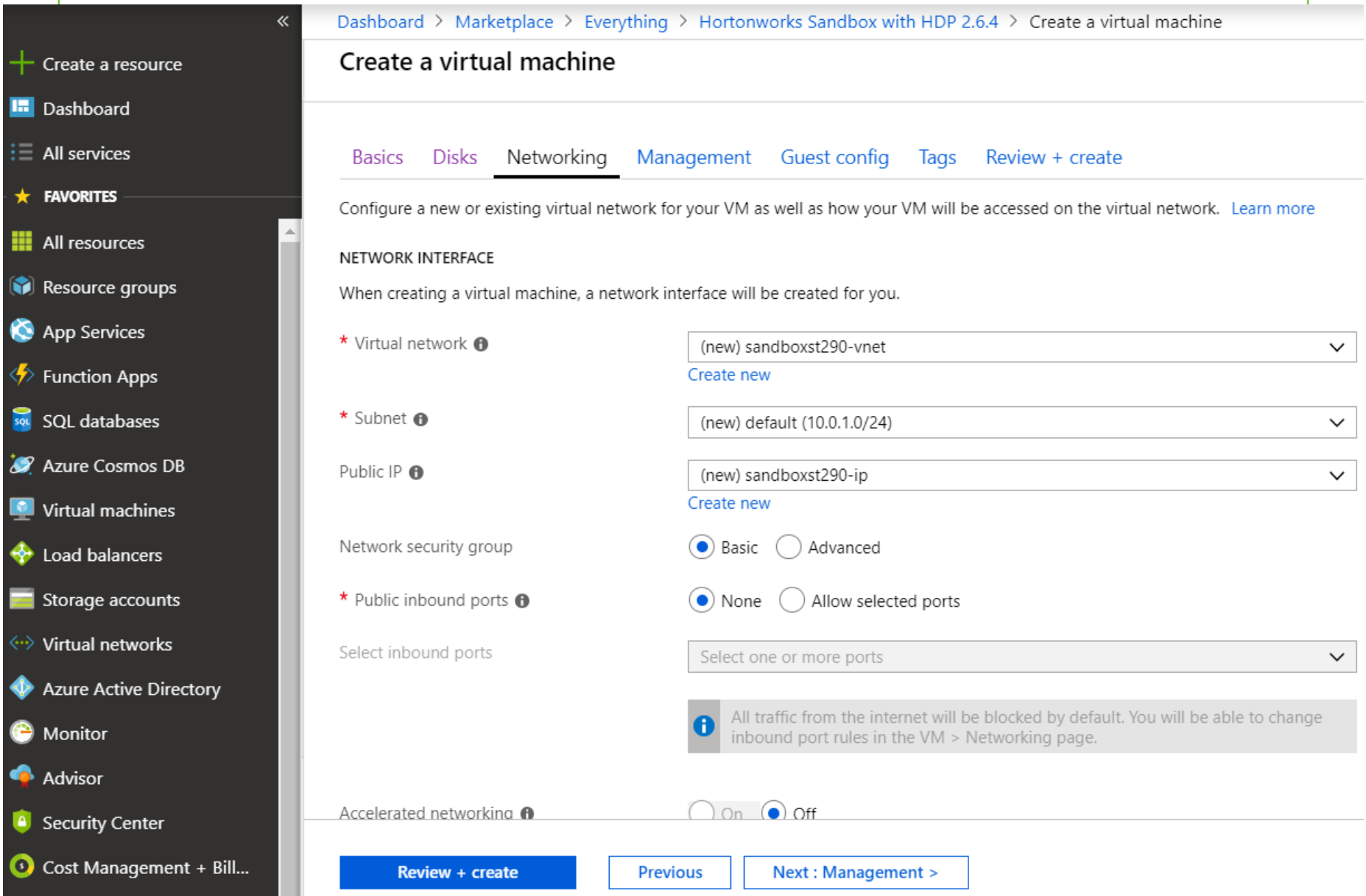
Select Password as the Authentication Type, suggested password: FRE7831@2019



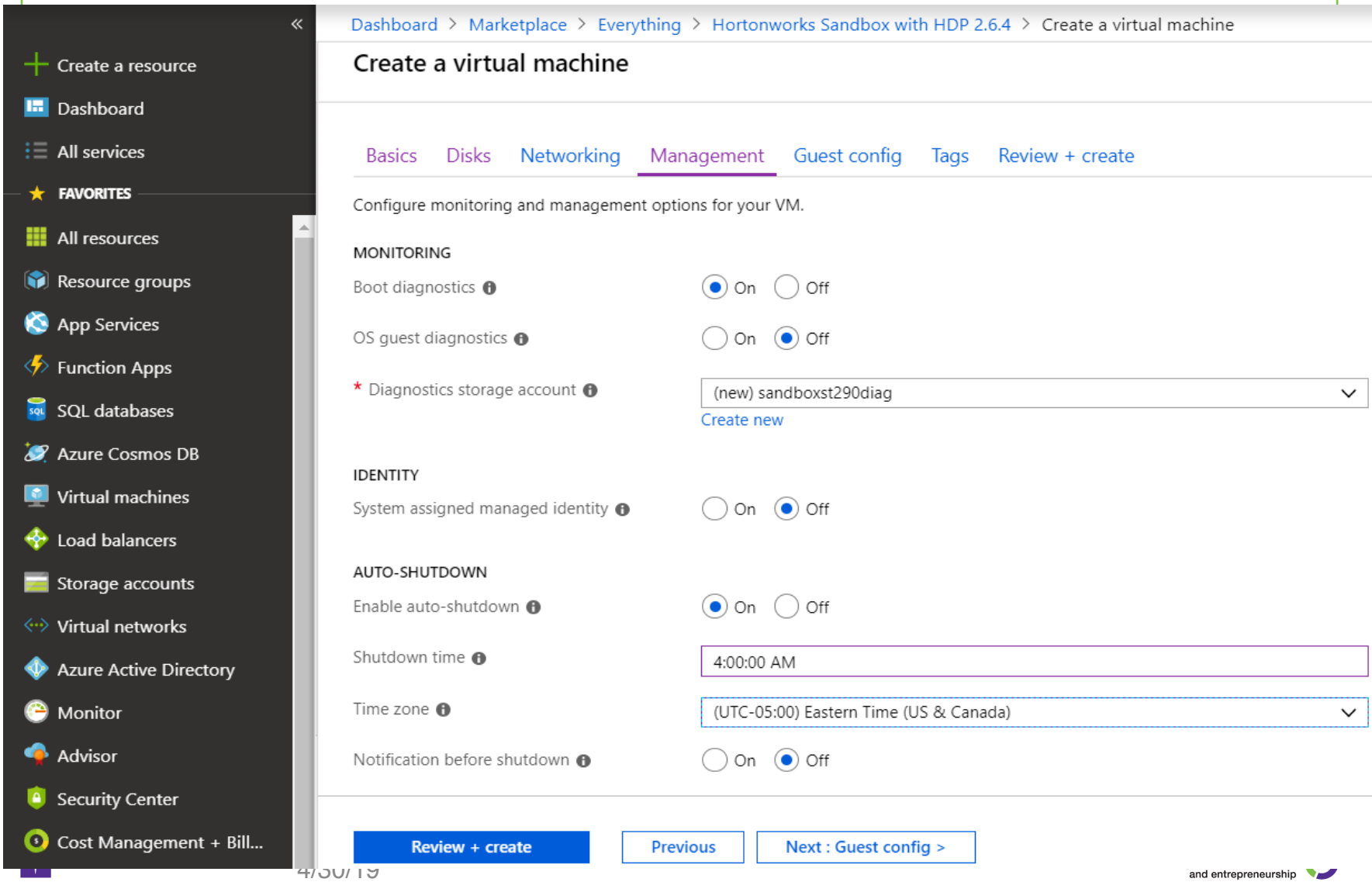
Select Virtual Machine Disk Type: Standard SSD

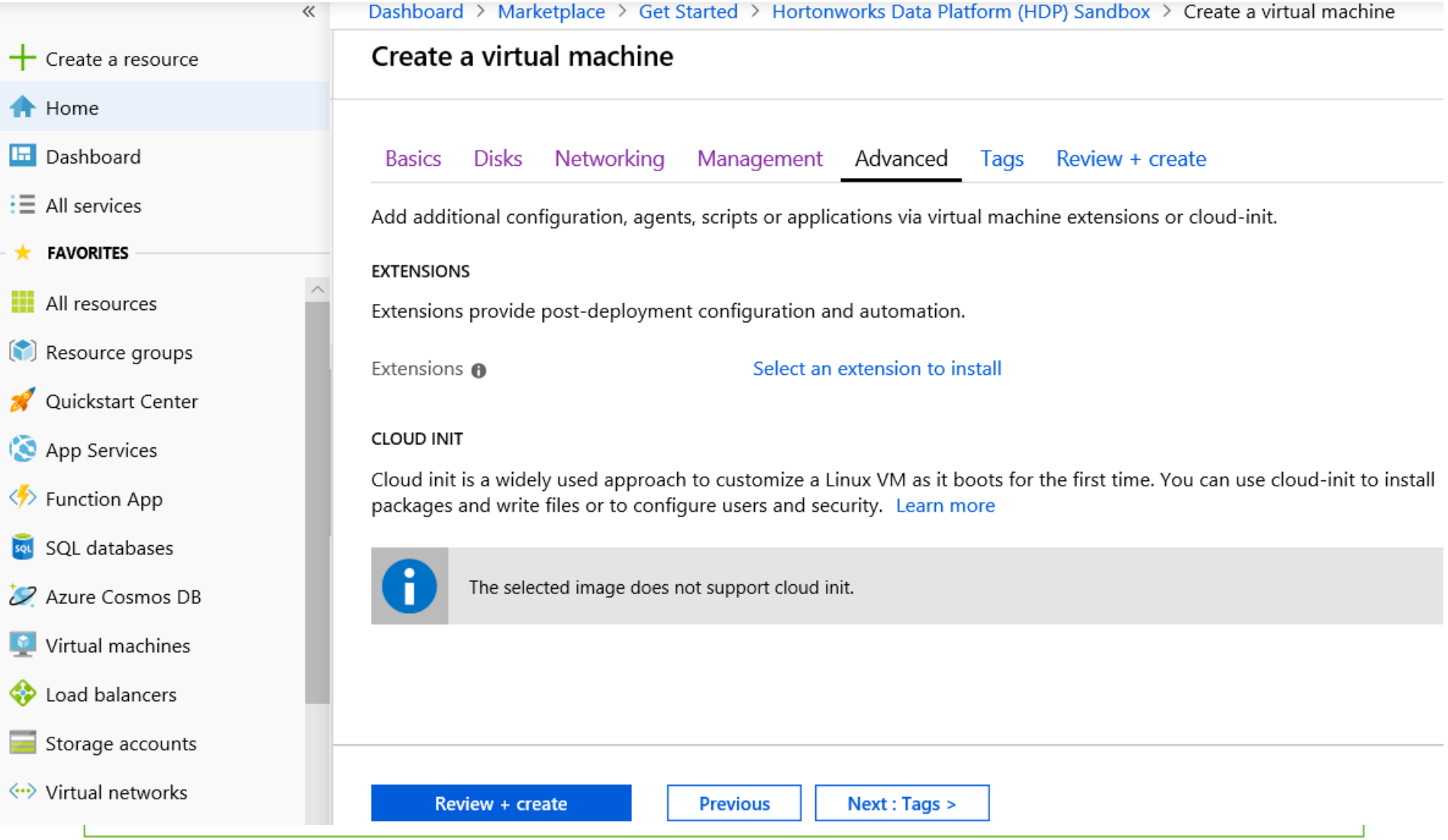


Verify your VM network setting similar to the following

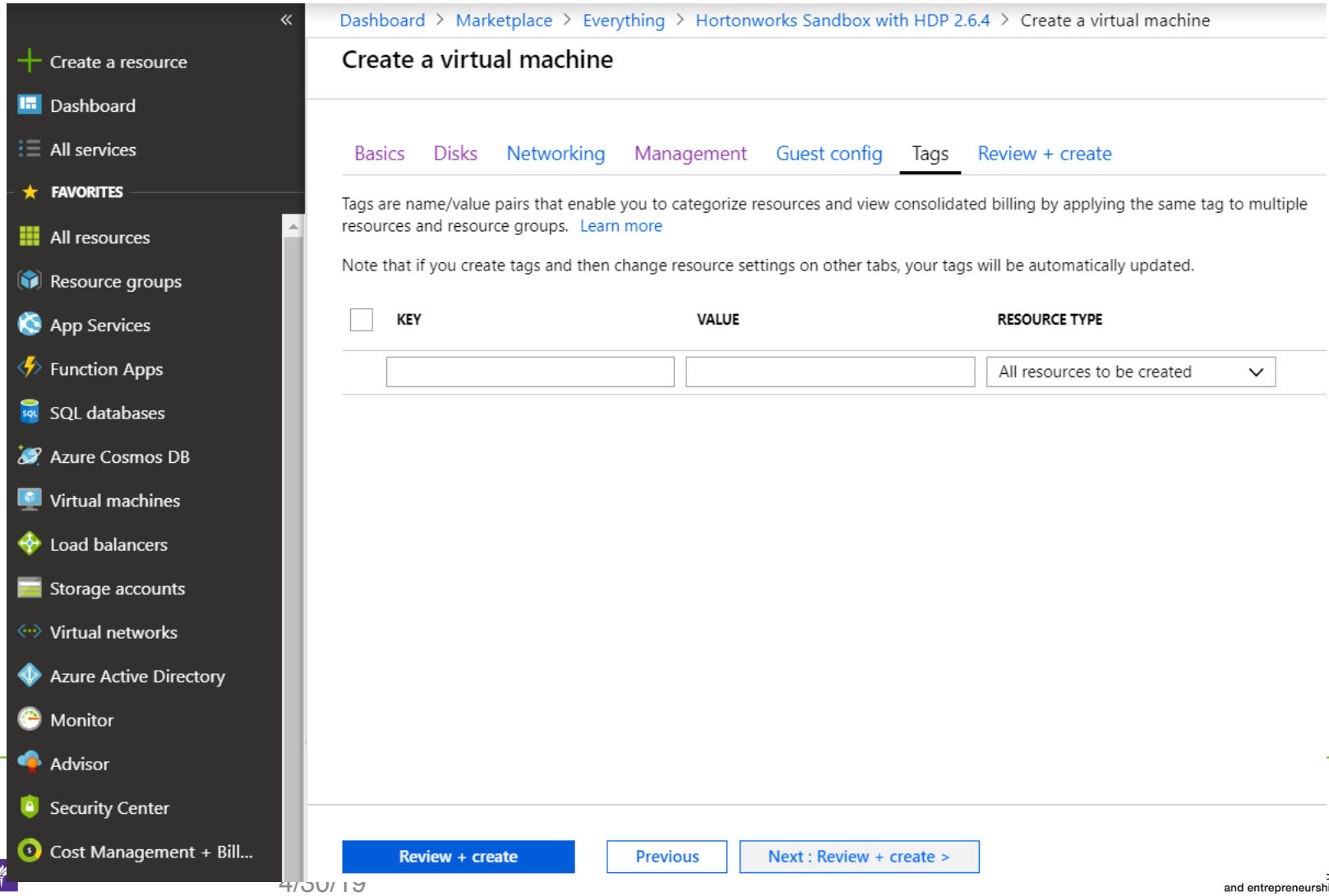


Verify your VM management setting similar to the following

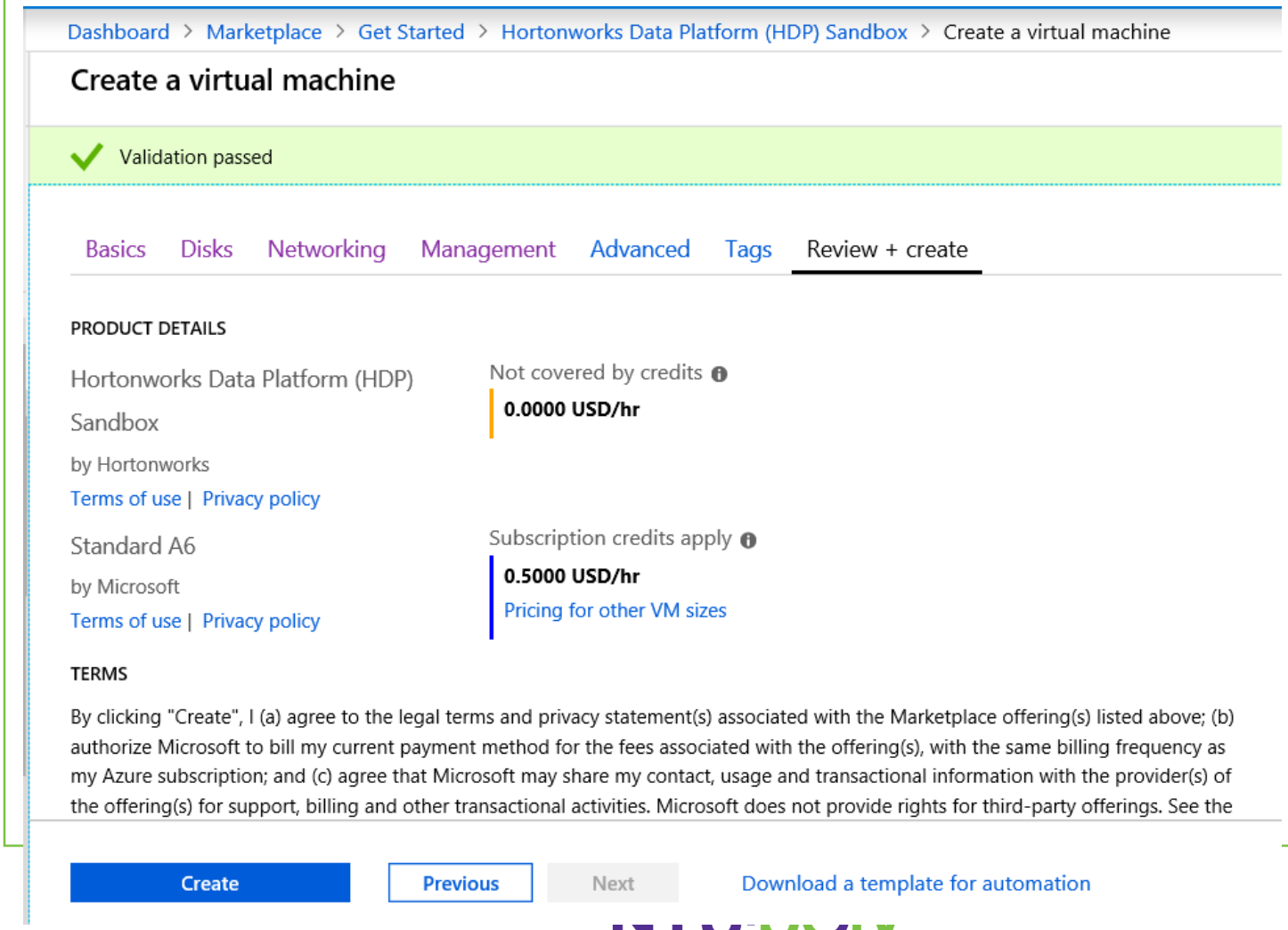



Create a resource
Home
Dashboard
All services
FAVORITES
All resources
Resource groups
Quickstart Center
App Services
Function App
SQL databases
Azure Cosmos DB
Virtual machines
Load balancers
Storage accounts
Virtual networks
Dashboard > Marketplace > Get Started > Hortonworks Data Platform (HDP) Sandbox > Create a virtual machine
Create a virtual machine
Basics
Disks
Networking
Management
Advanced
Tags
Review + create
Add additional configuration, agents, scripts or applications via virtual machine extensions or cloud-init.
EXTENSIONS
Extensions provide post-deployment configuration and automation.
Extensions
Select an extension to install
CLOUD INIT
Cloud init is a widely used approach to customize a Linux VM as it boots for the first time. You can use cloud-init to install packages and write files or to configure users and security. Learn more
The selected image does not support cloud init.
Review + create
Previous
Next : Tags >

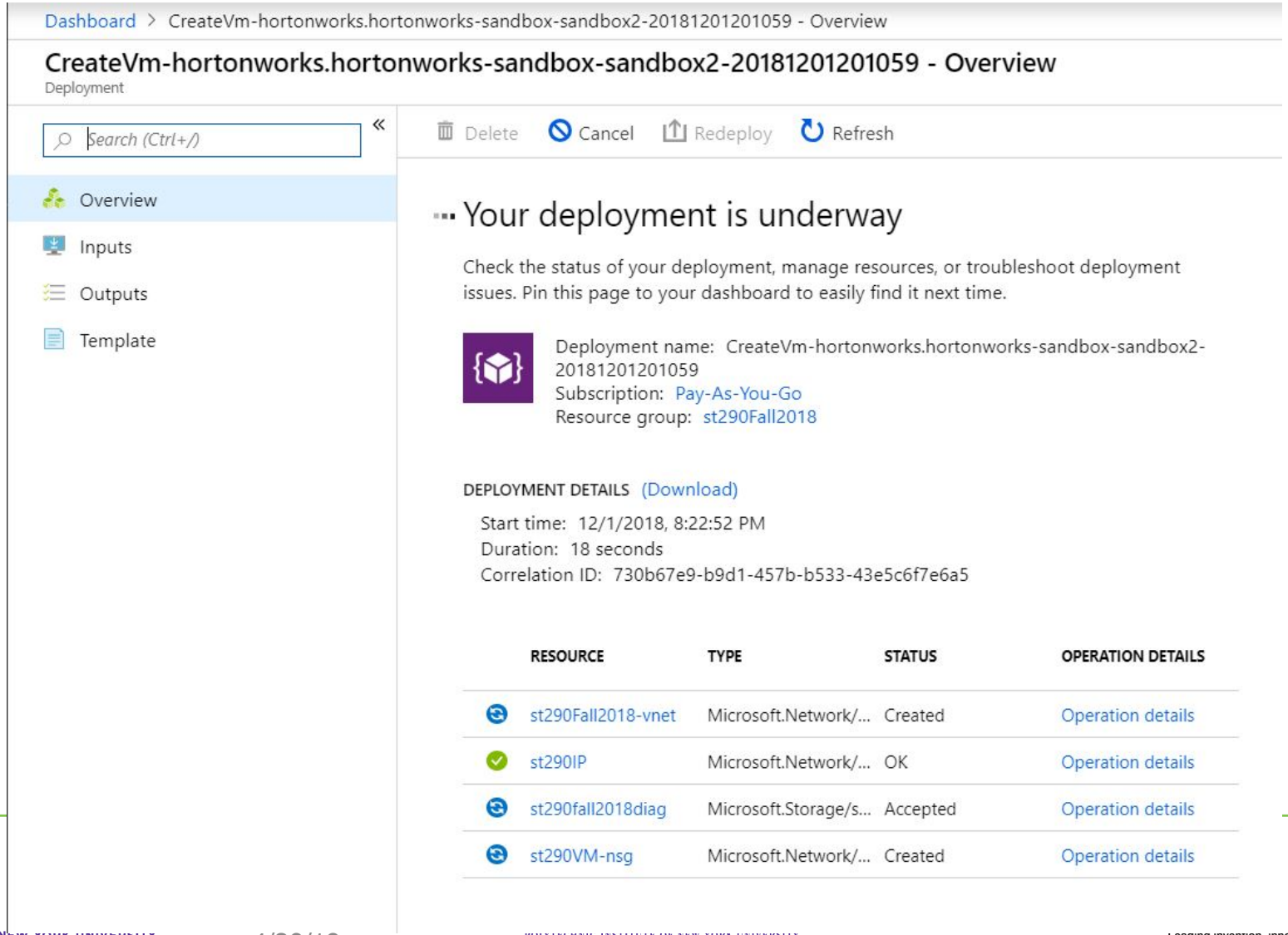
Verify your VM tags same as the following.
Continue to Review & Create your VM!



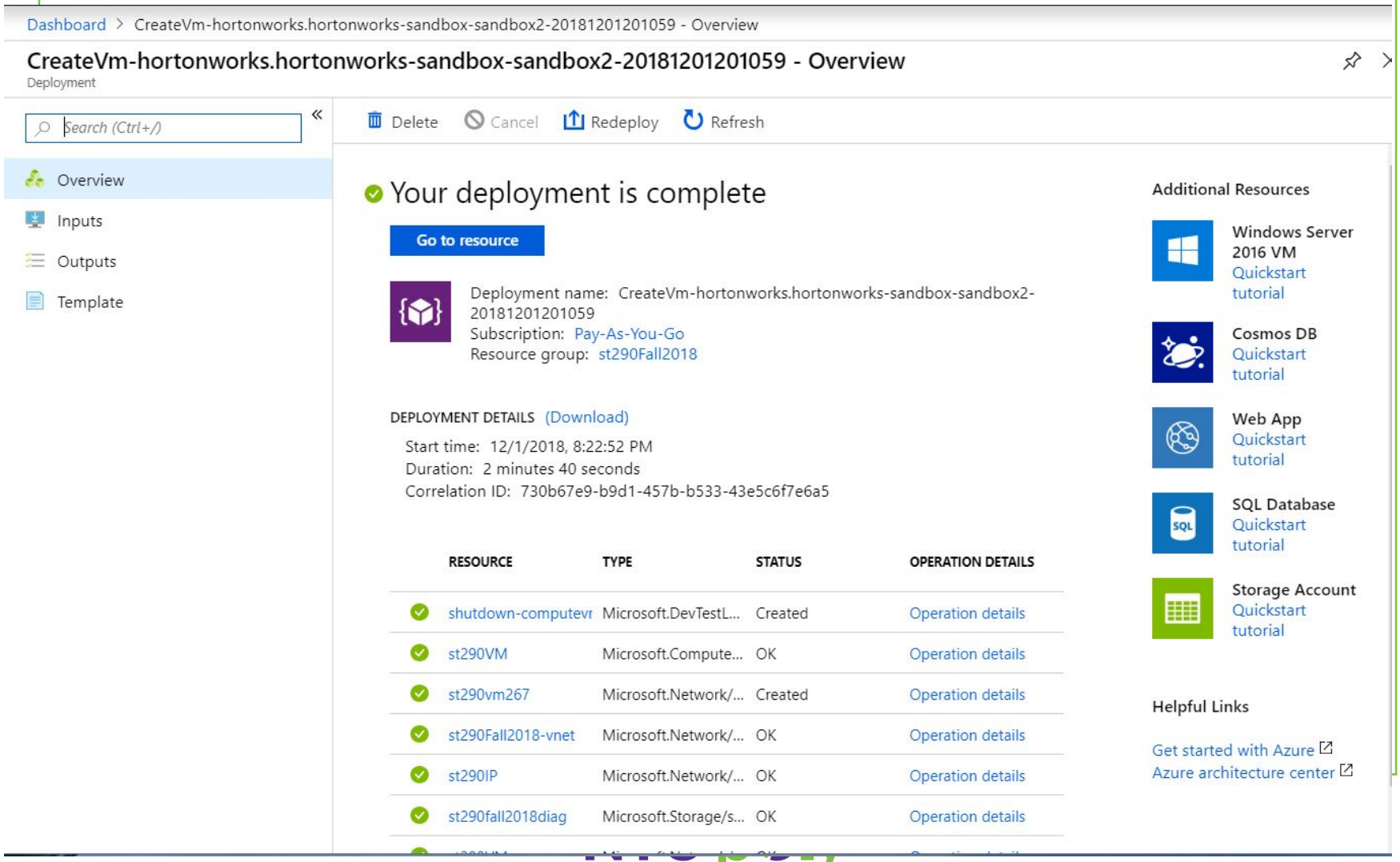
Wait for your virtual machine settings to be validated then click Create



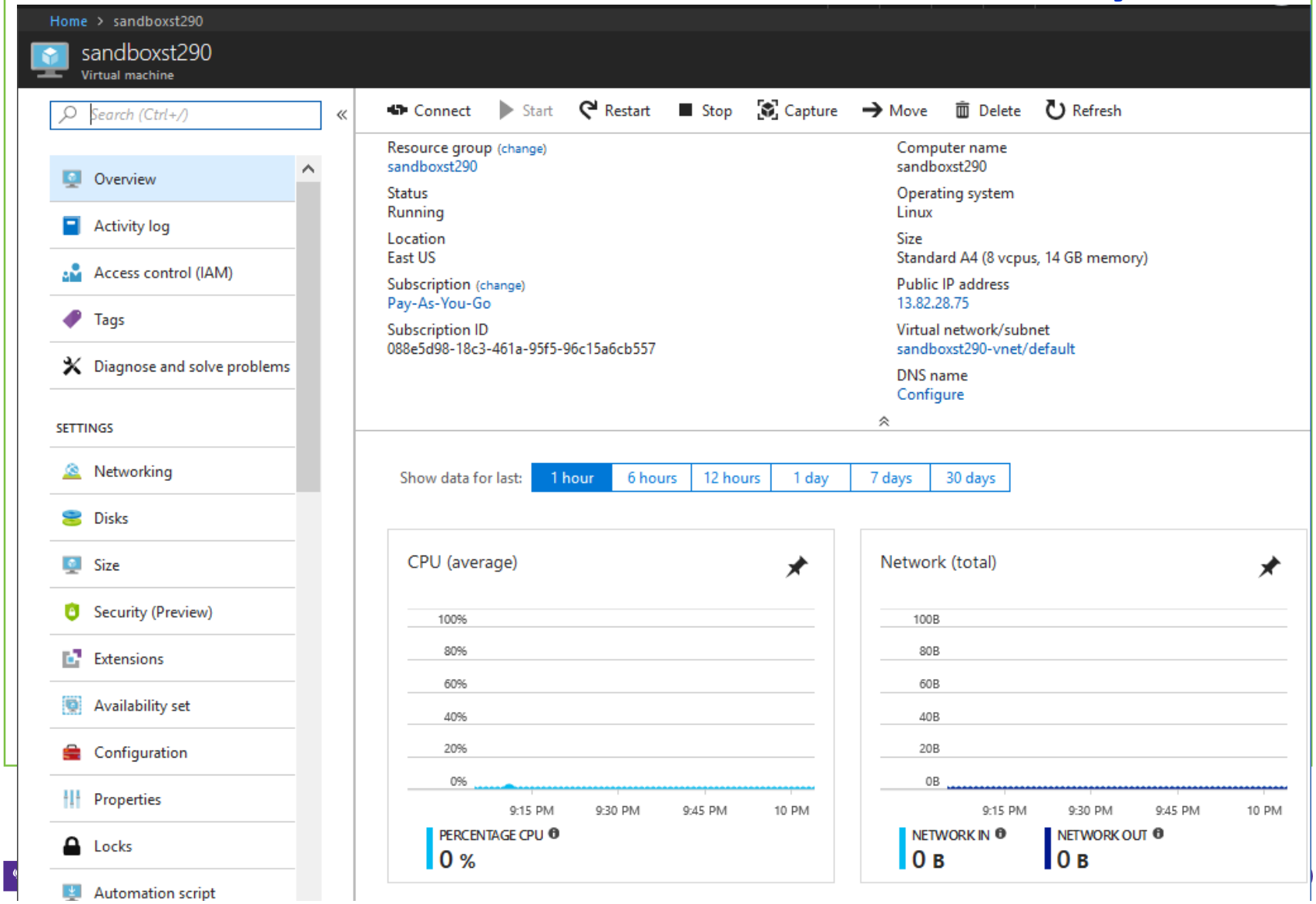
Your VM is in the process of deployment



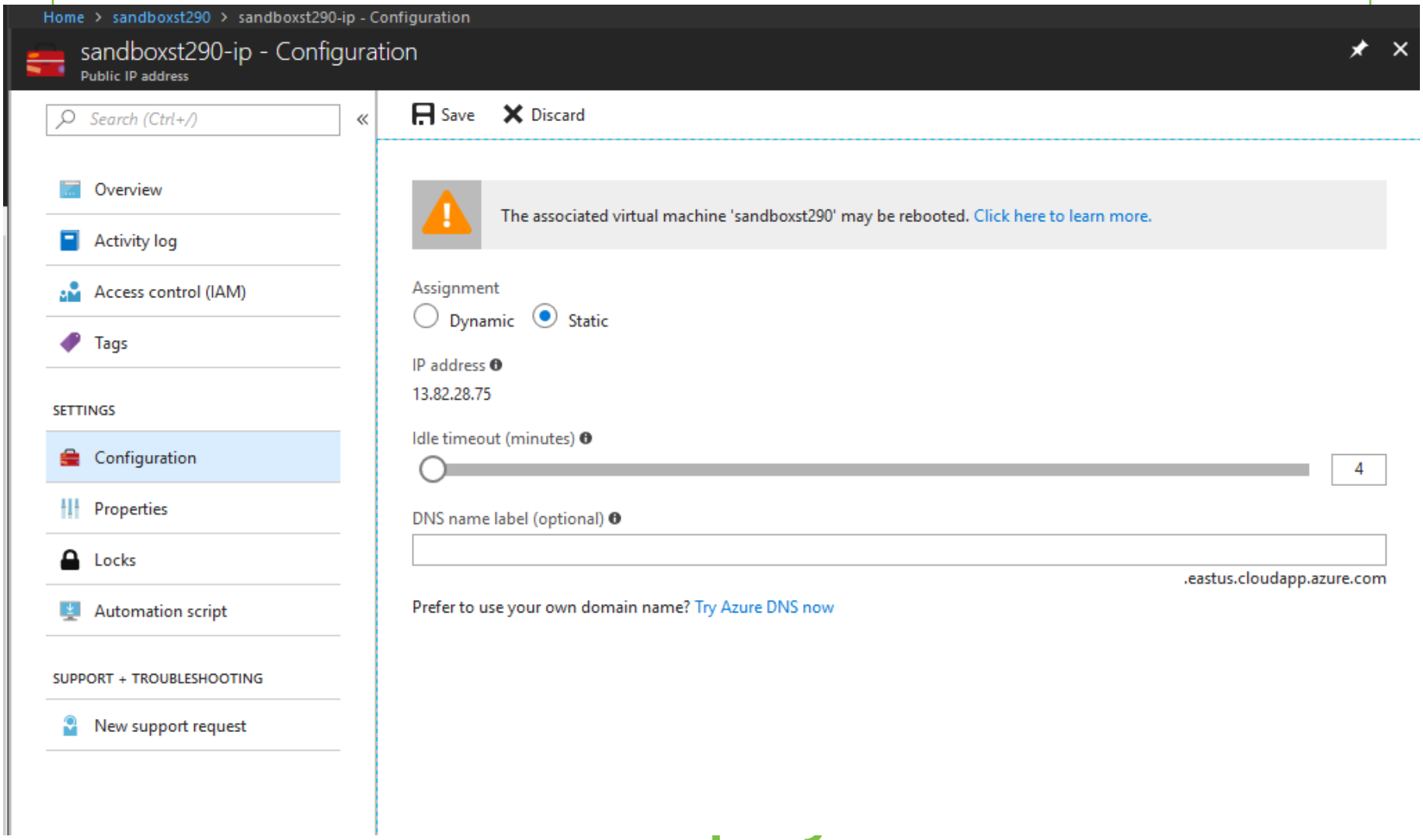
Wait until your VM deployment is complete



Your Microsoft Azure Virtual Machine is ready!

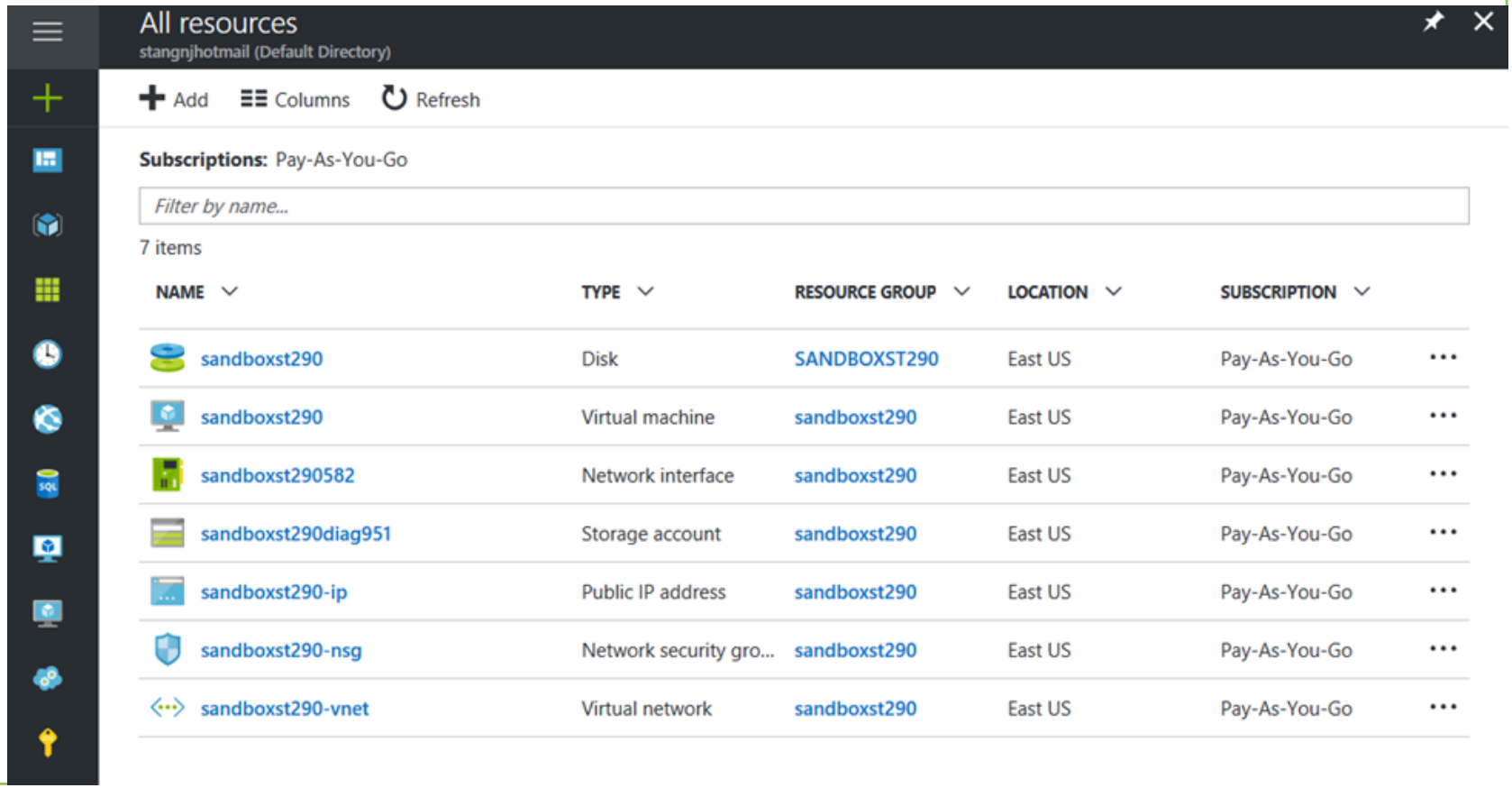


Change your Public IP address to Static (Click Save)

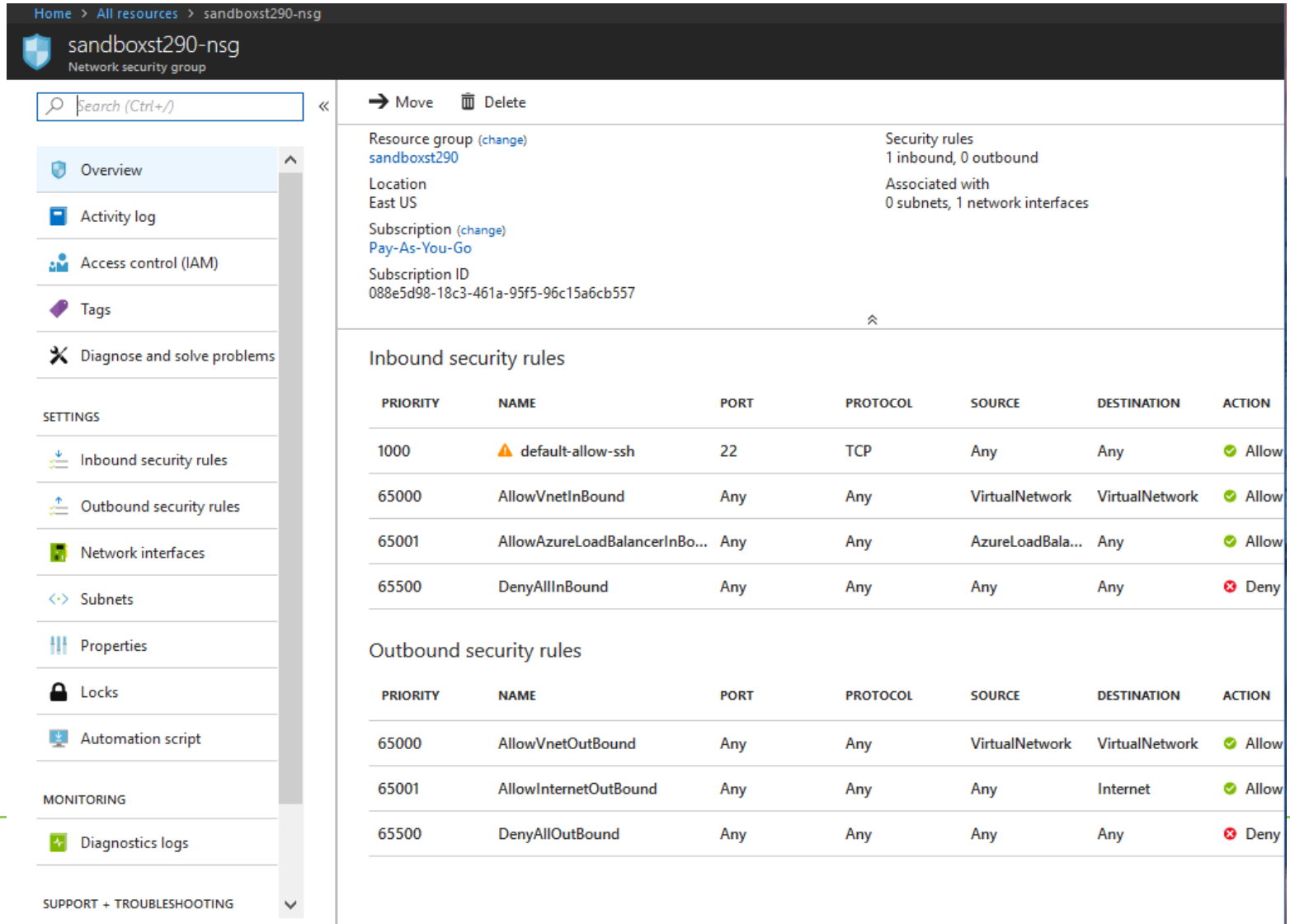


Open firewall for access your VM

- Click All resources and select Network Security Group

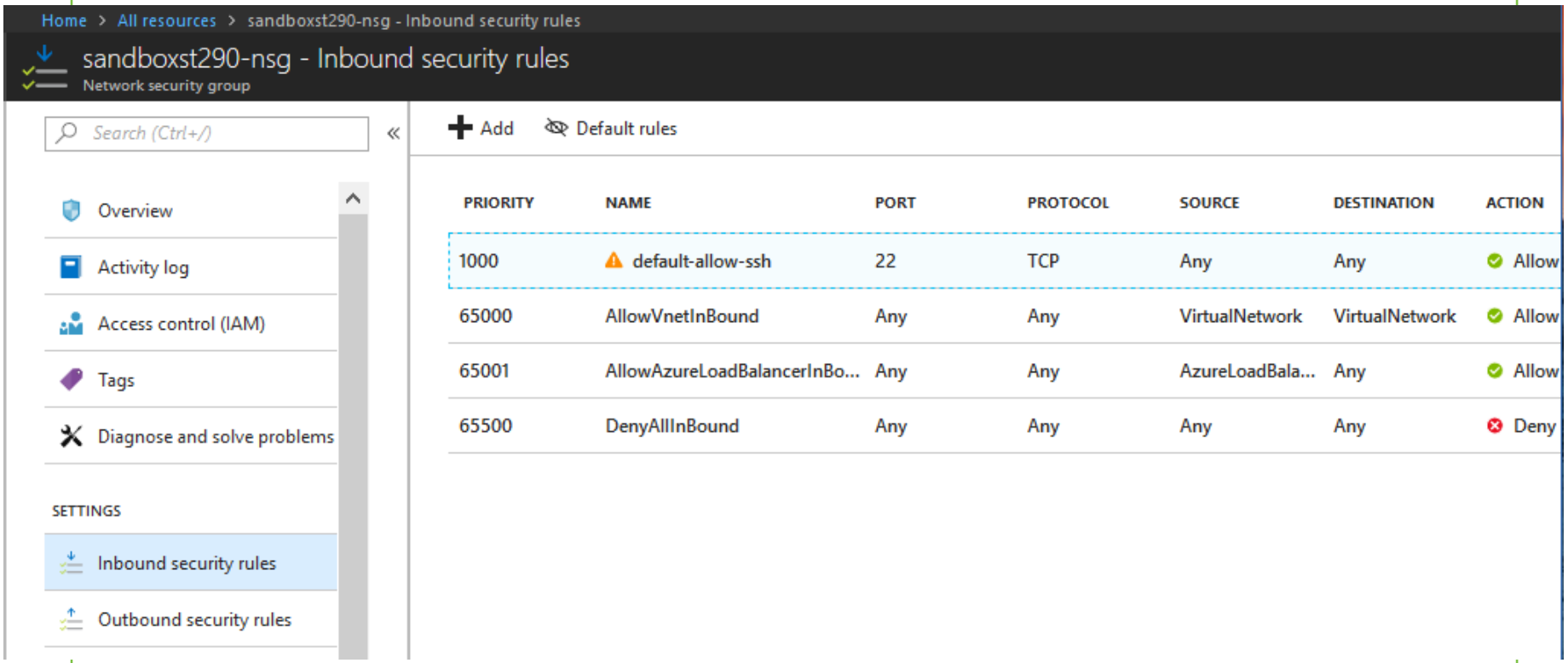


Select Network Security Group



Add an inbound rule to open firewall of your VM

- Click Inbound Security rules and then click Add



sandboxst290-nsg - Inbo

Network security group

Search (Ctrl+/)

- Overview
- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems

Settings

- Inbound security rules
- Outbound security rules
- Network interfaces
- Subnets
- Properties
- Locks
- Export template

Monitoring

- Diagnostic settings

Add inbound security rule

sandboxst290-nsg

Basic

* Source: Any

* Source port ranges: *

* Destination: Any

* Destination port ranges: 1024-65535

* Protocol: Any | TCP | UDP

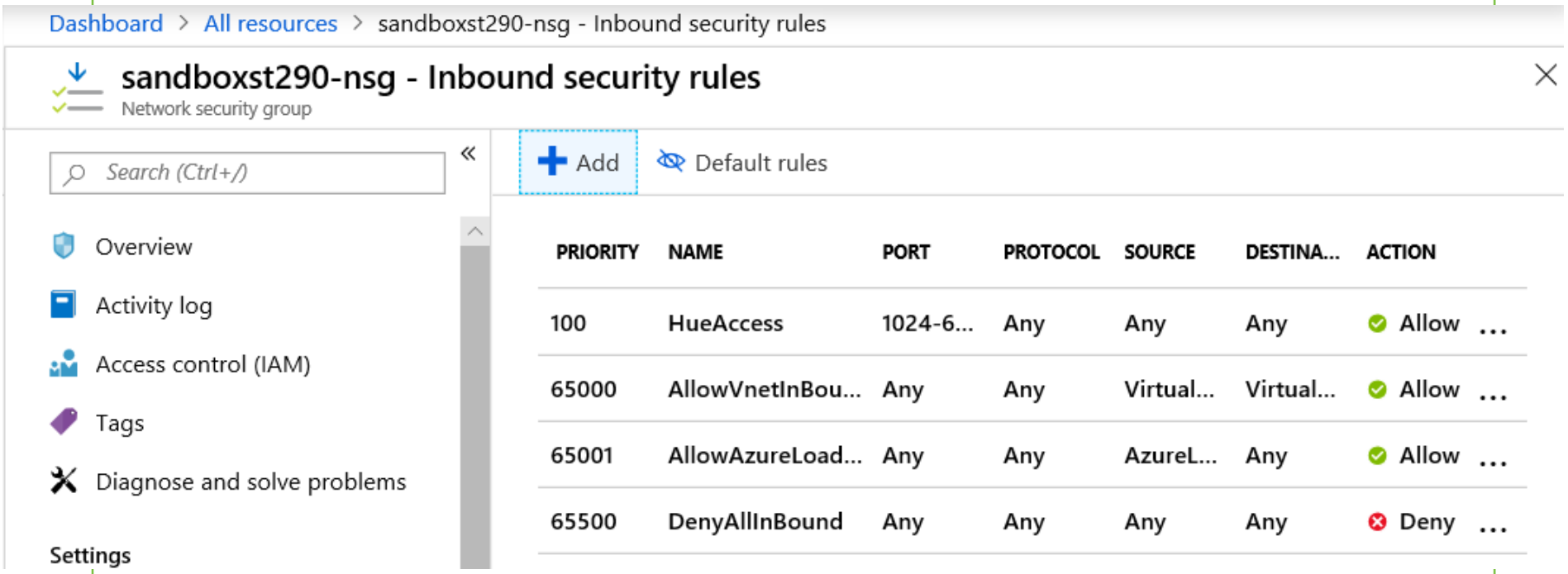
* Action: Allow | Deny

* Priority: 100

* Name: HueAccess

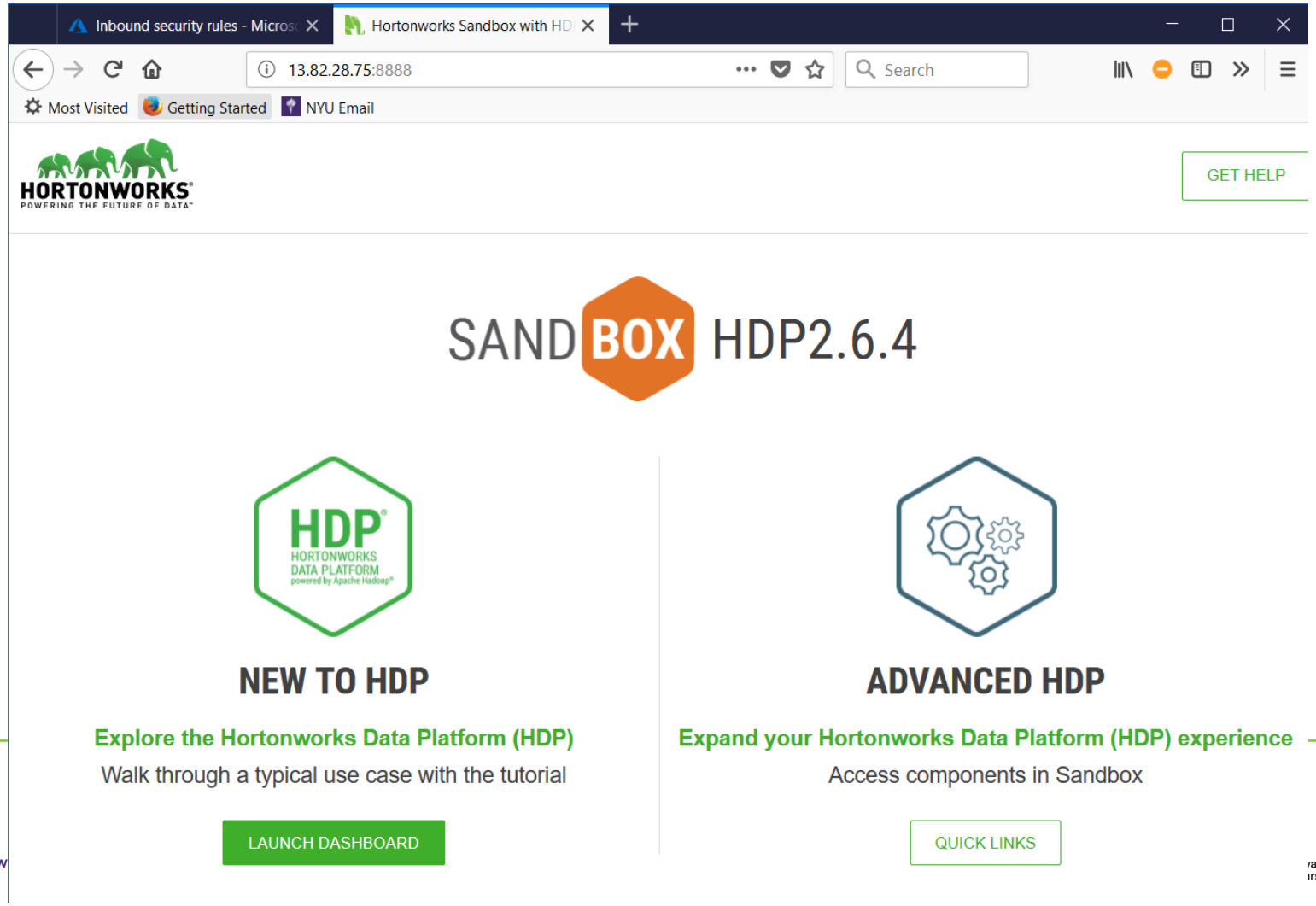
Add

After the inbound rule is created

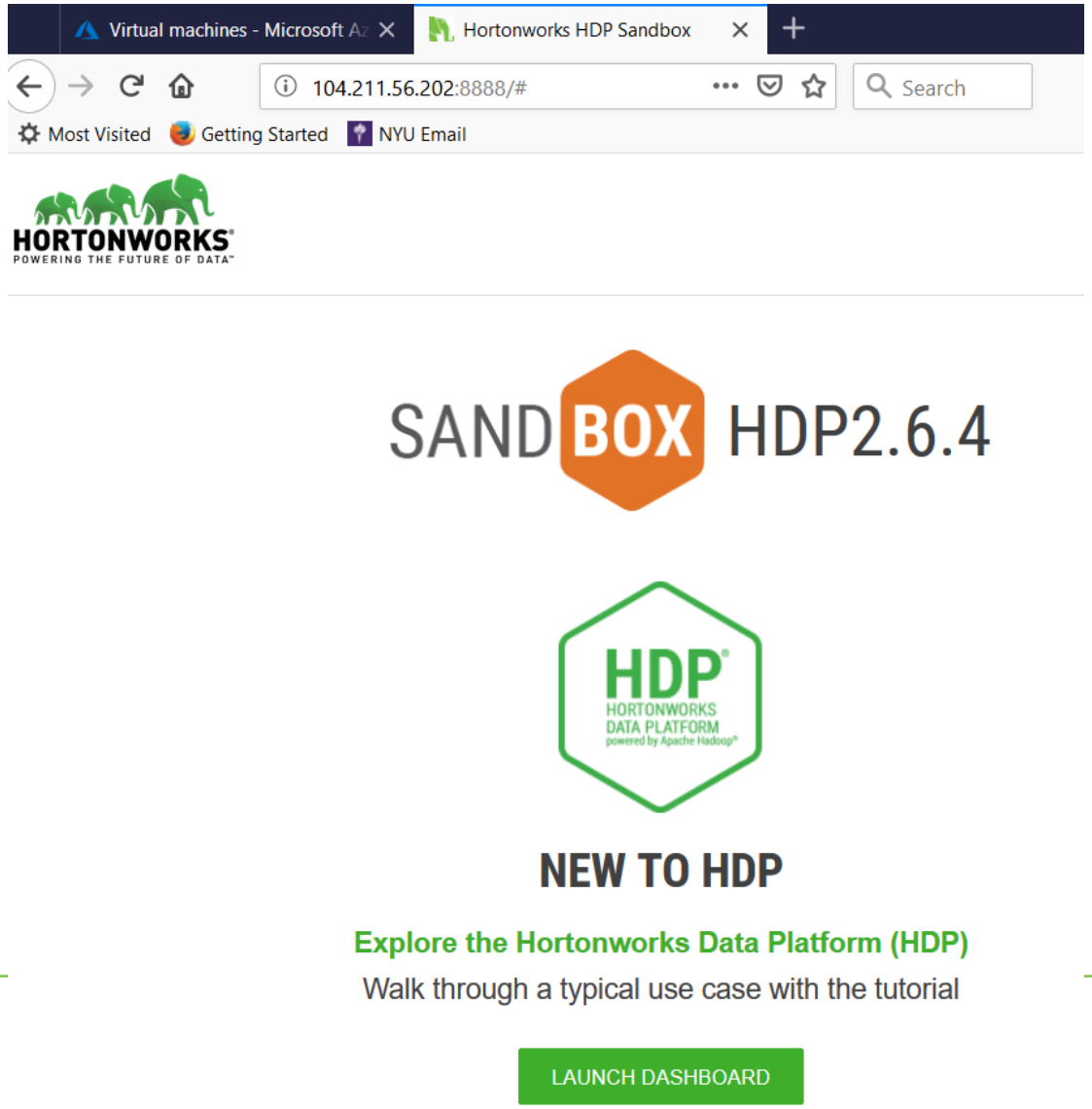


Hortonworks HDP 2.6.4 Portal

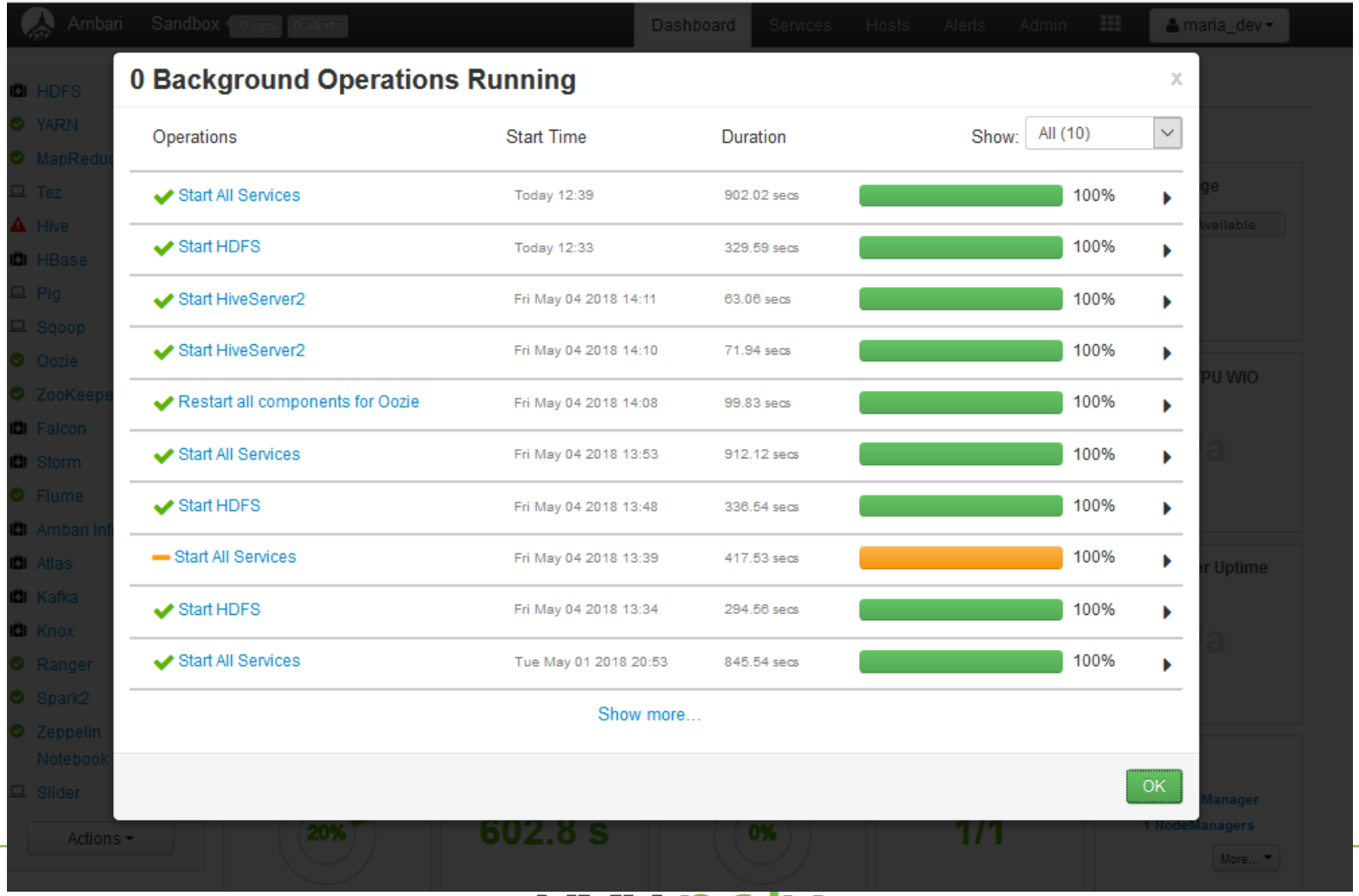
Wait a few mins (to let firewall rule take effect) and then type your IP address:8888 in your Web browser



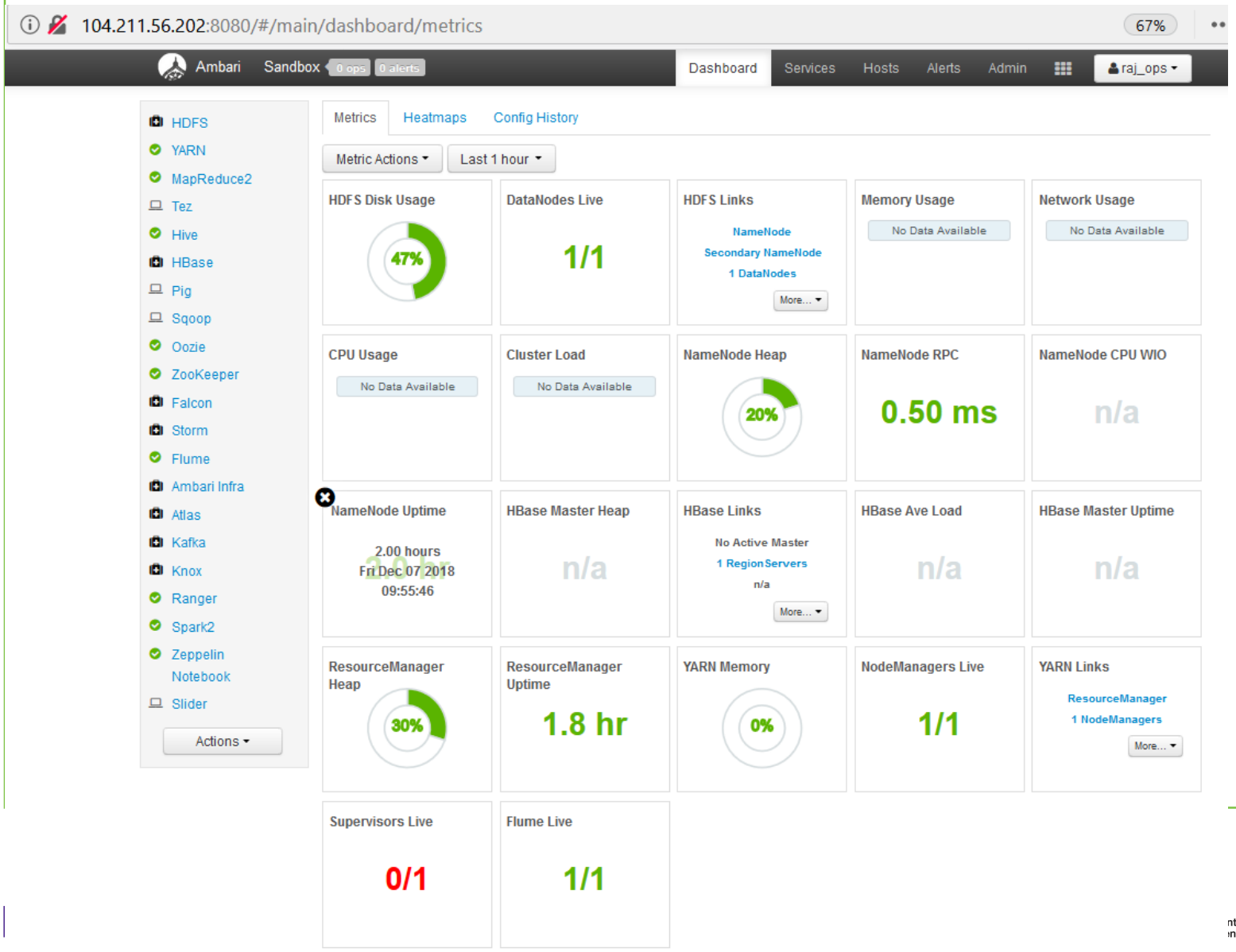
Launch Dashboard with Username/Password: raj_ops/raj_ops



Make sure all the processes are running

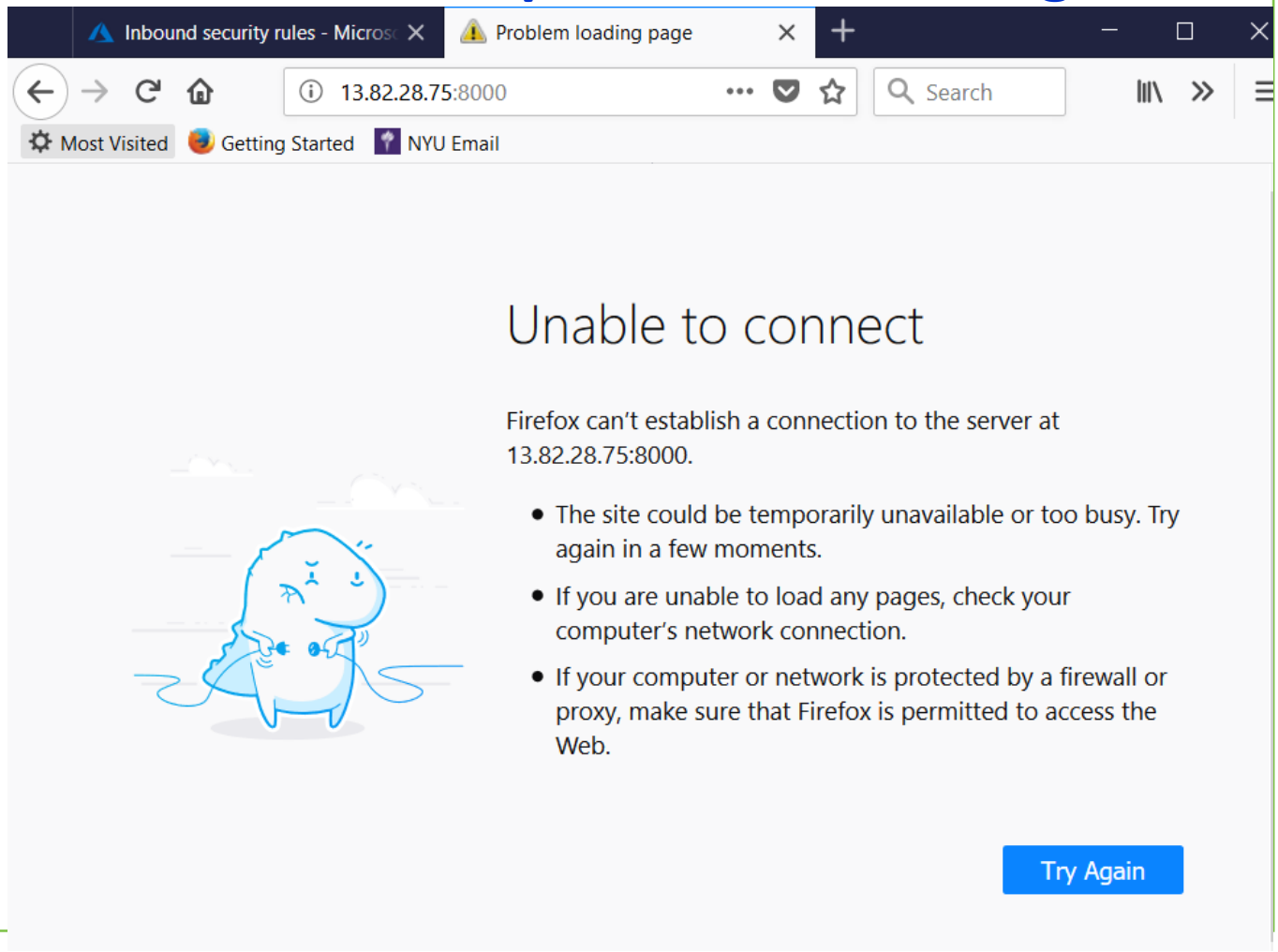


Make sure all the processes are running (except Supervisors Line)

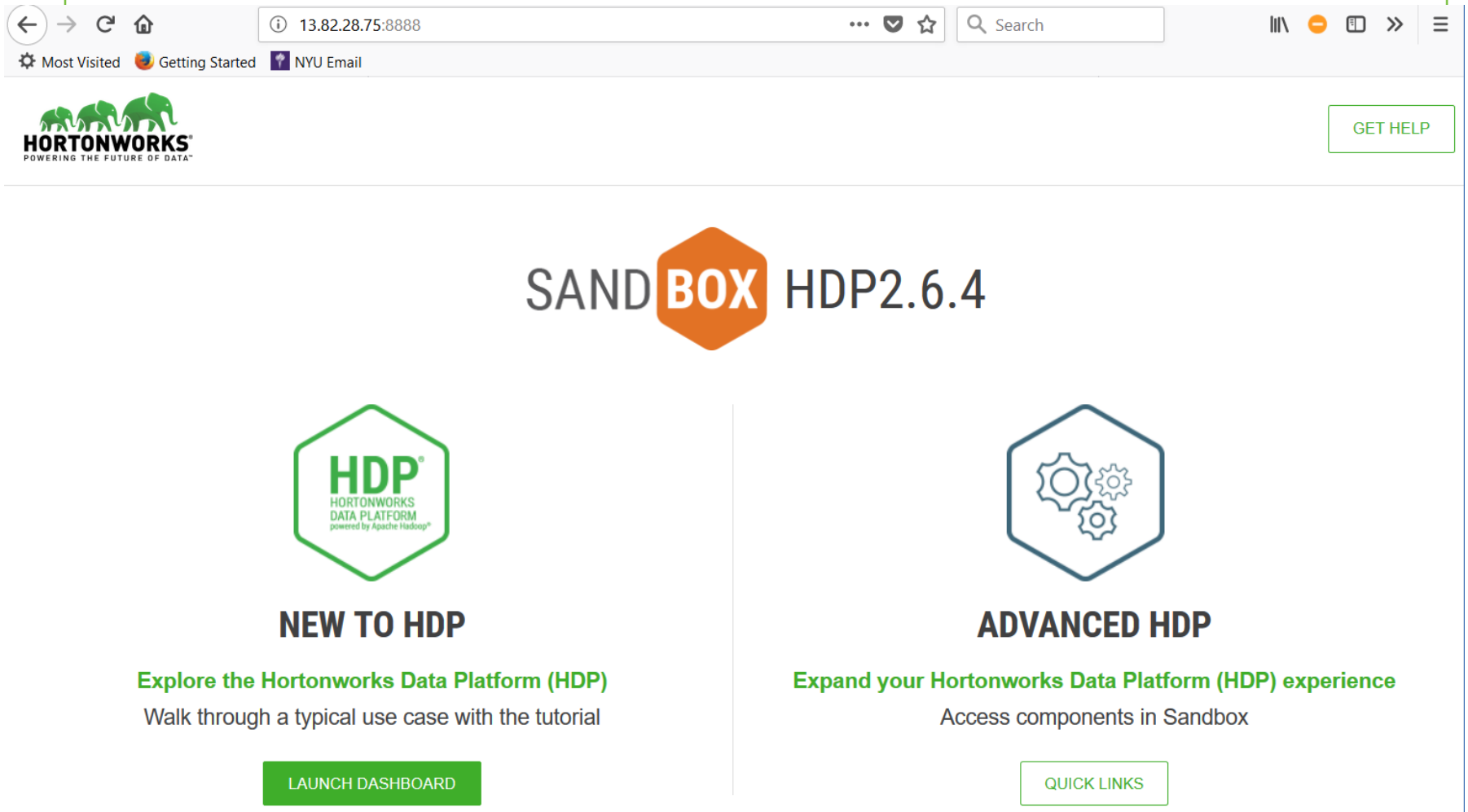


However Azure HDP defect prevents launching HUE!

We will use SSH client to access the Hortonworks virtual machine in Microsoft Azure Cloud and fix this defect.



Click Quick Links for Advanced HDP



Click SSH Client

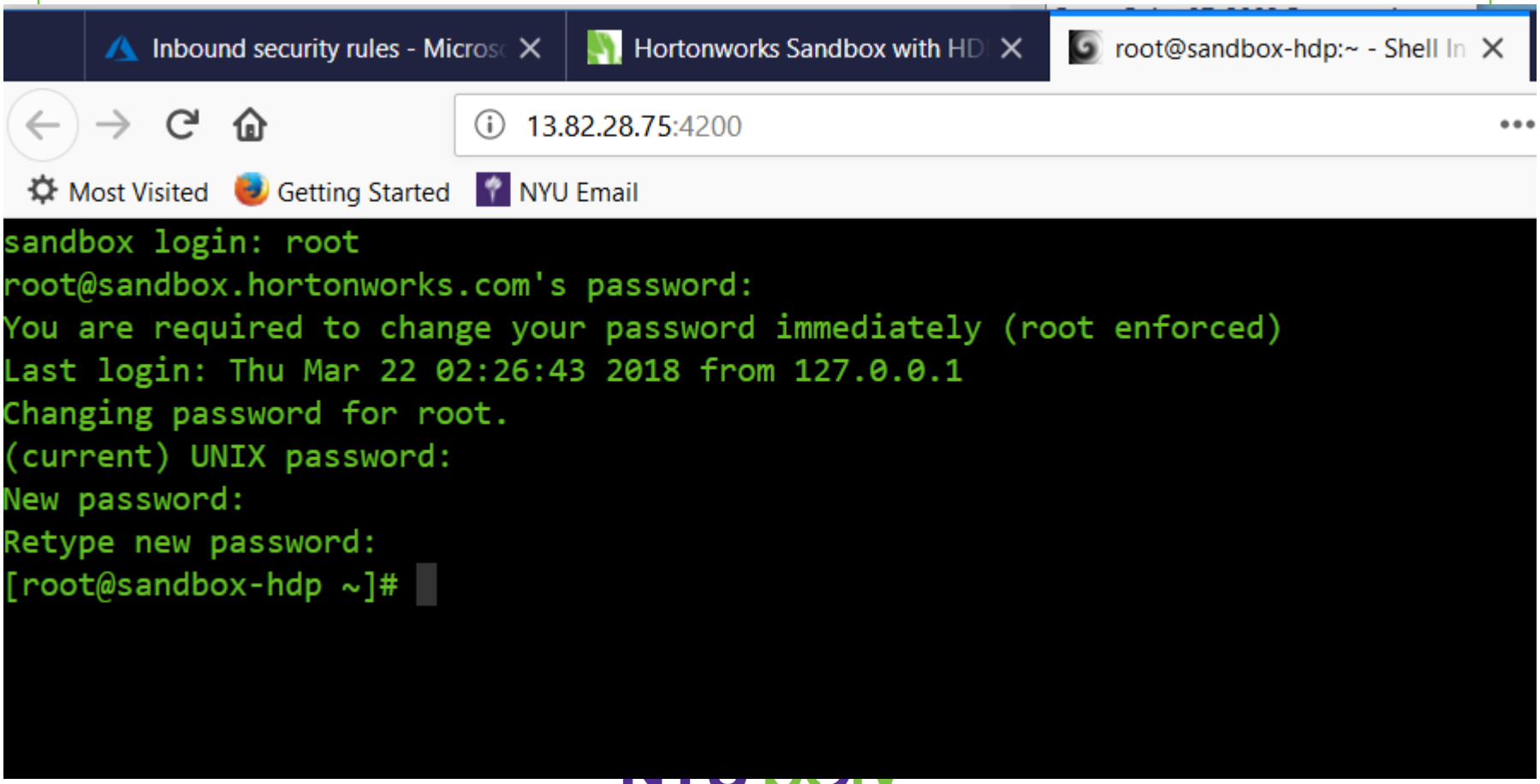
GET HELP

ADVANCED HDP QUICK LINKS

AMBARI	ATLAS	WORKFLOW MANAGER
RANGER	ZEPPELIN	SSH CLIENT

Click http://YourIPAddress:4200

- User name is root, password: hadoop. Change password to FRE7831@2019

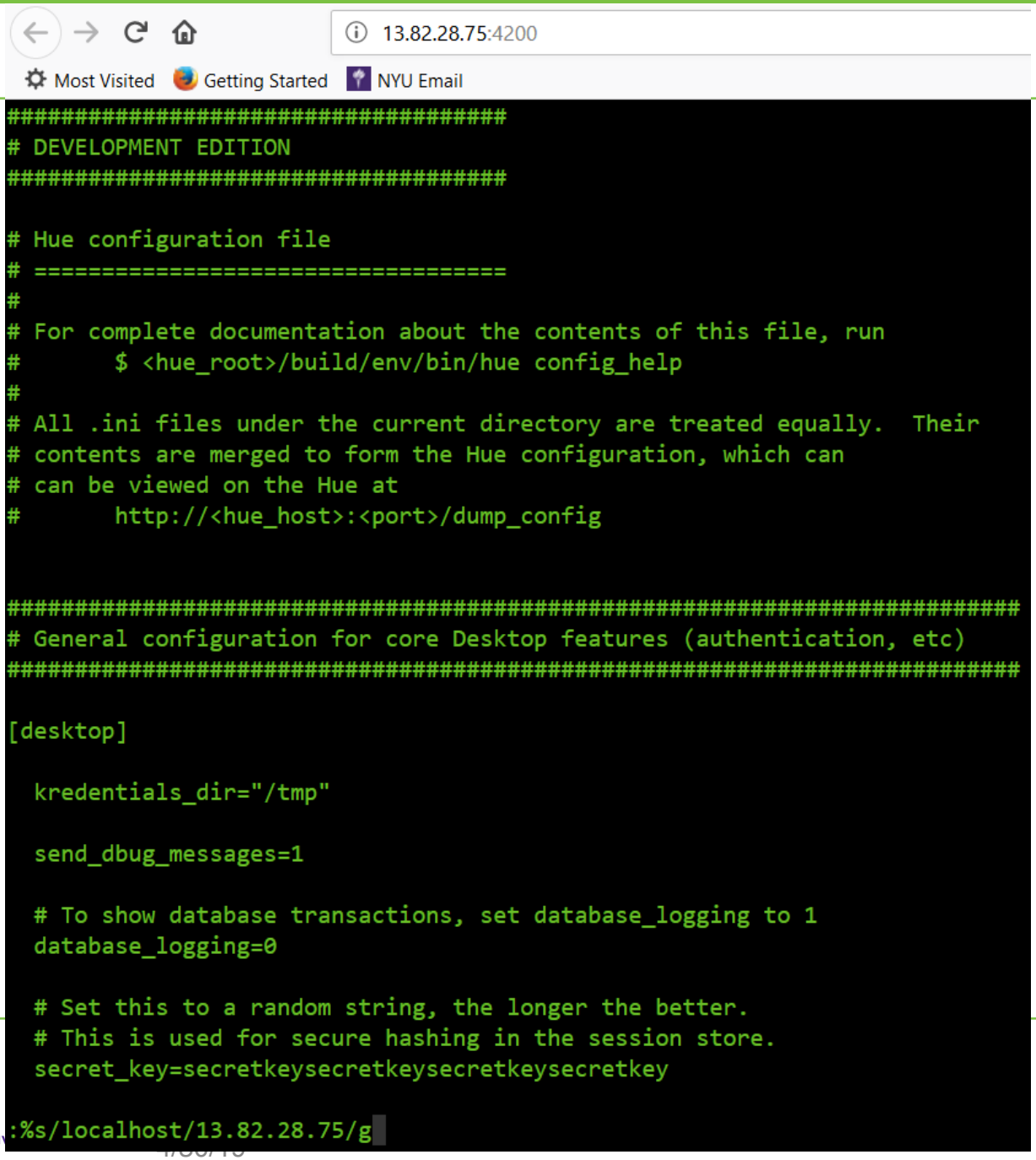


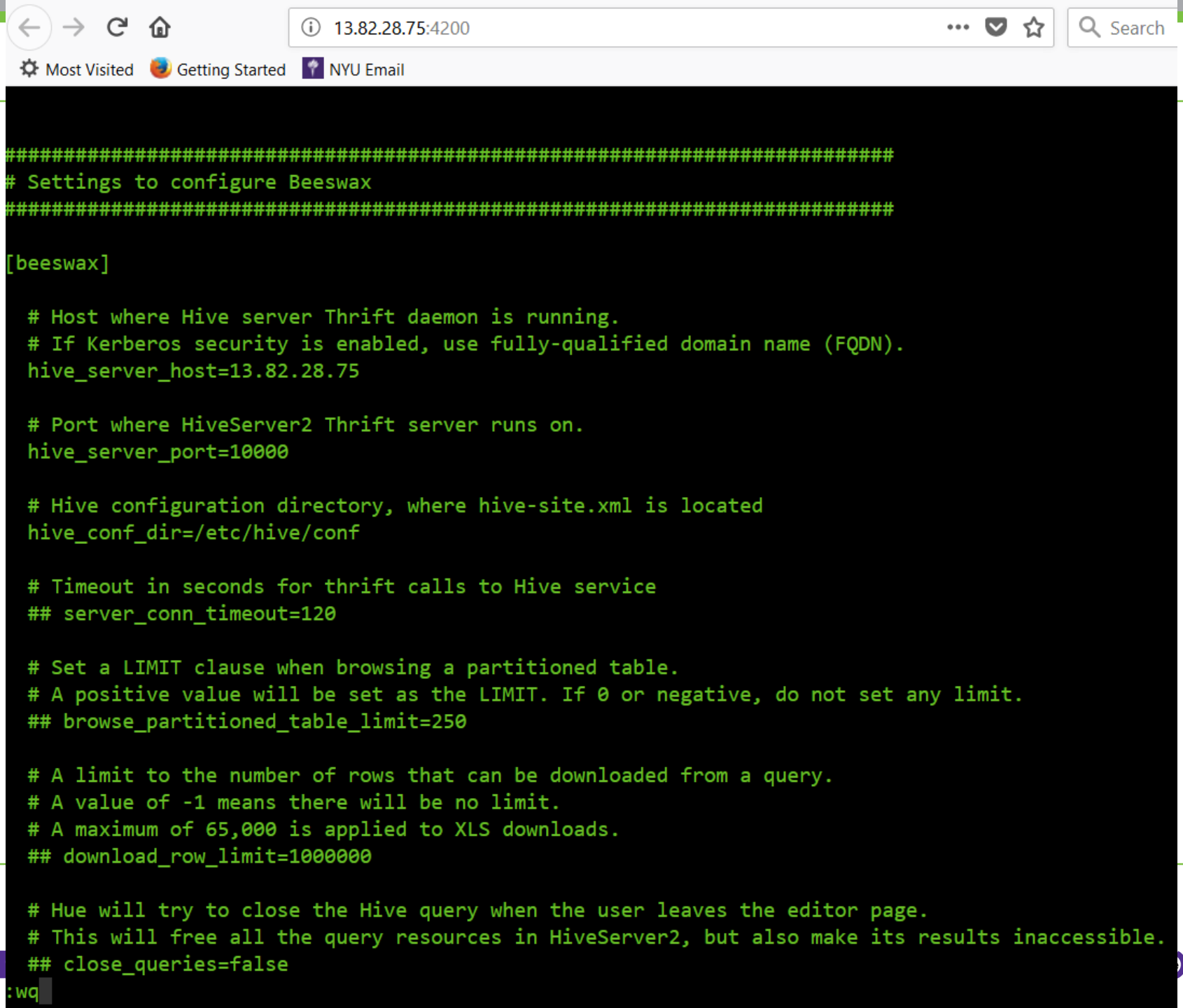
Steps for Fix Azure HDP bug

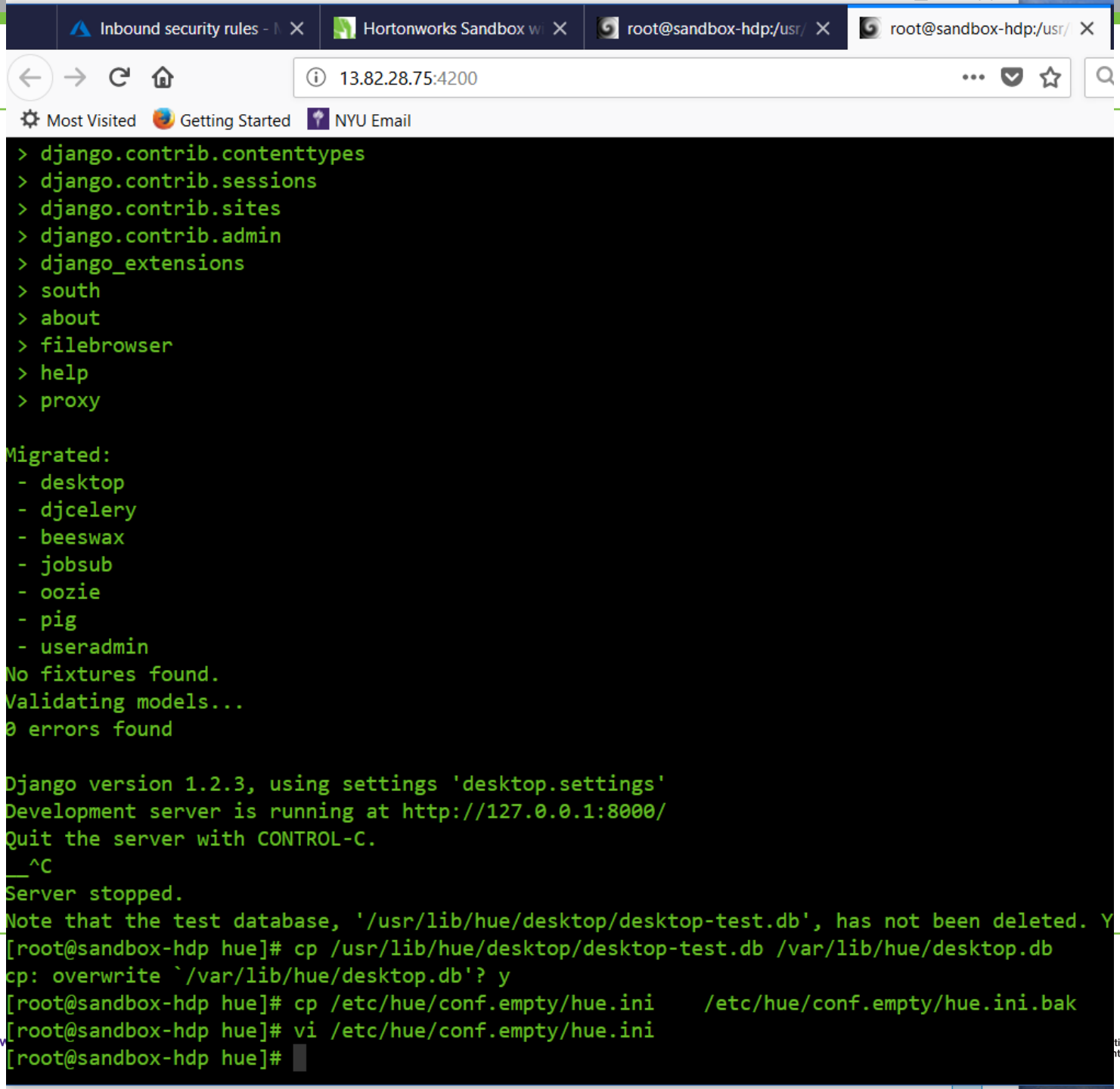
1. Change the owner of /usr/lib/hue to hue:hue:
   Command: **chown -R hue:hue /usr/lib/hue**
2. Replace the corrupted database.
   1) create a test server
   Command: **/usr/lib/hue/build/env/bin/hue testserver**
   2) shutdown the server
      Command**: control + c**
   3) replace database
   Command: **cp /usr/lib/hue/desktop/desktop-test.db /var/lib/hue/desktop.db**
3. Change the configuration of hue
   command: **cp /etc/hue/conf.empty/hue.ini /etc/hue/conf.empty/hue.ini.bak**
   command**: vi /etc/hue/conf.empty/hue.ini**
   1) change all localhost to virtual machine ip address
      command in the vi/vim: **%s/localhost/<ip address>/g**
   command: **shift : type w, press enter**
   2) uncomment the line to link hive conf directory in the beeswax section.(line 441)
   command: 1. **:441**
   2. **uncomment it**
   3. **shift : type wq, press enter**
4. Restart the hue service: Command: **service hue start**

```
[root@sandbox-hdp hue]# ls -l
total 76
-rw-r--r-- 1 hue  hue   1704 Feb  1 10:34 app.reg
drwxr-xr-x 1 hue  hue   4096 Feb  1 10:34 apps
drwxr-xr-x 1 hue  hue   4096 Feb  1 10:34 build
drwxr-xr-x 1 hue  hue   4096 Feb  1 10:34 desktop
drwxr-xr-x 1 hue  hue   4096 Feb  1 10:34 ext
-rw-r--r-- 1 hue  hue  11358 Jan  4 10:26 LICENSE.txt
drwxr-xr-x 1 hue  hue   4096 Feb  1 10:41 logs
-rw-r--r-- 1 hue  hue   8170 Jan  4 10:26 Makefile
-rw-r--r-- 1 root root    44 Jan  4 10:31 Makefile.buildvars
-rw-r--r-- 1 root root  8505 Jan  4 10:26 Makefile.sdk
-rw-r--r-- 1 root root  3652 Jan  4 10:26 Makefile.vars
-rw-r--r-- 1 root root  2302 Jan  4 10:26 Makefile.vars.priv
drwxr-xr-x 1 hue  hue   4096 Feb  1 10:34 tools
-rw-r--r-- 1 hue  hue     16 Jan  4 10:26 VERSION
[root@sandbox-hdp hue]# chown hue:hue Makefile*
[root@sandbox-hdp hue]# ll
total 76
-rw-r--r-- 1 hue hue  1704 Feb  1 10:34 app.reg
drwxr-xr-x 1 hue hue  4096 Feb  1 10:34 apps
drwxr-xr-x 1 hue hue  4096 Feb  1 10:34 build
drwxr-xr-x 1 hue hue  4096 Feb  1 10:34 desktop
drwxr-xr-x 1 hue hue  4096 Feb  1 10:34 ext
-rw-r--r-- 1 hue hue 11358 Jan  4 10:26 LICENSE.txt
drwxr-xr-x 1 hue hue  4096 Feb  1 10:41 logs
-rw-r--r-- 1 hue hue  8170 Jan  4 10:26 Makefile
-rw-r--r-- 1 hue hue    44 Jan  4 10:31 Makefile.buildvars
-rw-r--r-- 1 hue hue  8505 Jan  4 10:26 Makefile.sdk
-rw-r--r-- 1 hue hue  3652 Jan  4 10:26 Makefile.vars
-rw-r--r-- 1 hue hue  2302 Jan  4 10:26 Makefile.vars.priv
drwxr-xr-x 1 hue hue  4096 Feb  1 10:34 tools
-rw-r--r-- 1 hue hue    16 Jan  4 10:26 VERSION
[root@sandbox-hdp hue]#
```

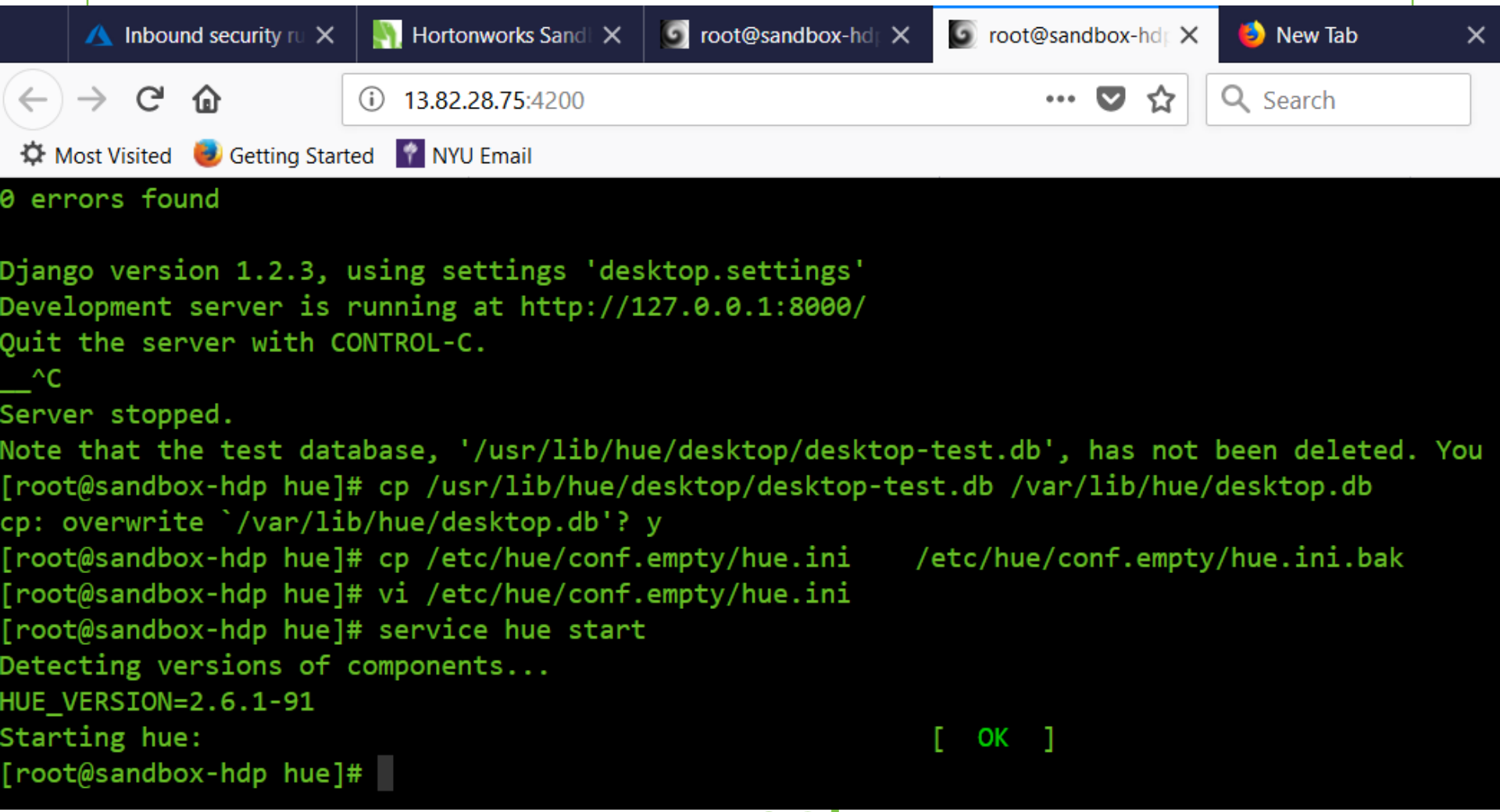
```
 > useradmin:0002_add_ldap_support
 - Migration 'useradmin:0002_add_ldap_support' is marked for no-dry-run.
 - Loading initial data for useradmin.
No fixtures found.

Synced:
 > django.contrib.auth
 > django.contrib.contenttypes
 > django.contrib.sessions
 > django.contrib.sites
 > django.contrib.admin
 > django_extensions
 > south
 > about
 > filebrowser
 > help
 > proxy

Migrated:
 - desktop
 - djcelery
 - beeswax
 - jobsub
 - oozie
 - pig
 - useradmin
No fixtures found.
Validating models...
0 errors found

Django version 1.2.3, using settings 'desktop.settings'
Development server is running at http://127.0.0.1:8000/
Quit the server with CONTROL-C.
```


13.82.28.75:4200
Most Visited
Getting Started
NYU Email
#####################################
DEVELOPMENT EDITION
#####################################
Hue configuration file
===================================
#
For complete documentation about the contents of this file, run
$ <hue_root>/build/env/bin/hue config_help
#
All .ini files under the current directory are treated equally. Their
contents are merged to form the Hue configuration, which can
can be viewed on the Hue at
http://<hue_host>:<port>/dump_config
###
General configuration for core Desktop features (authentication, etc)
###
[desktop]
kredentials_dir="/tmp"
send_dbug_messages=1
# To show database transactions, set database_logging to 1
database_logging=0
# Set this to a random string, the longer the better.
# This is used for secure hashing in the session store.
secret_key=secretkeysecretkeysecretkeysecretkey
:%s/localhost/13.82.28.75/g

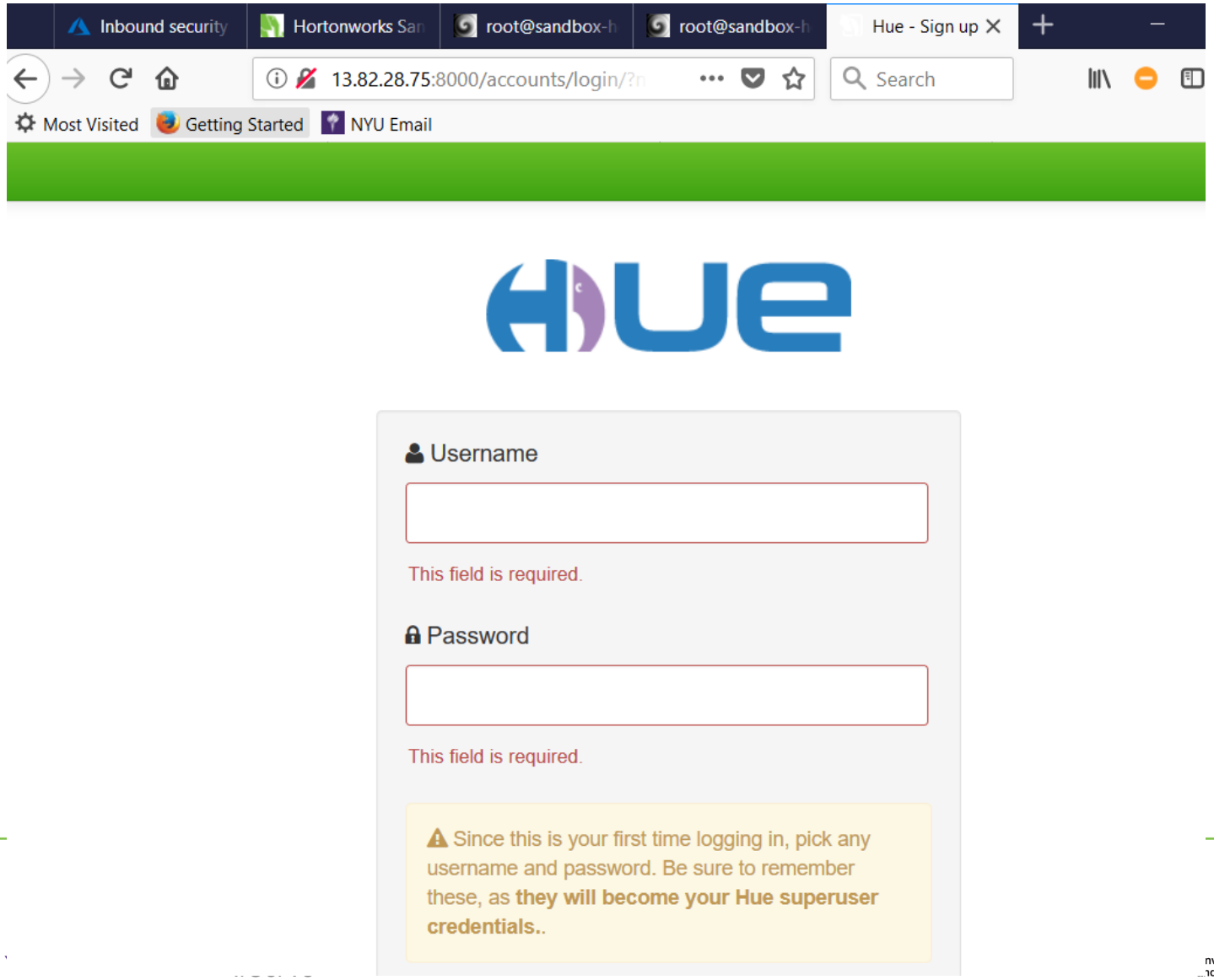

13.82.28.75:4200
Search
Most Visited
Getting Started
NYU Email
##
Settings to configure Beeswax
##
[beeswax]
# Host where Hive server Thrift daemon is running.
# If Kerberos security is enabled, use fully-qualified domain name (FQDN).
hive_server_host=13.82.28.75
# Port where HiveServer2 Thrift server runs on.
hive_server_port=10000
# Hive configuration directory, where hive-site.xml is located
hive_conf_dir=/etc/hive/conf
# Timeout in seconds for thrift calls to Hive service
server_conn_timeout=120
Set a LIMIT clause when browsing a partitioned table.
A positive value will be set as the LIMIT. If 0 or negative, do not set any limit.
browse_partitioned_table_limit=250
A limit to the number of rows that can be downloaded from a query.
A value of -1 means there will be no limit.
A maximum of 65,000 is applied to XLS downloads.
download_row_limit=1000000
Hue will try to close the Hive query when the user leaves the editor page.
This will free all the query resources in HiveServer2, but also make its results inaccessible.
close_queries=false
:wq

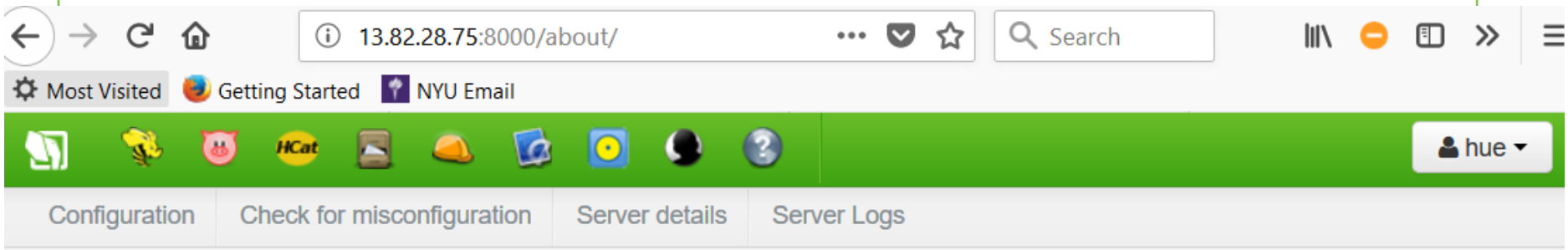

Inbound security rules -
Hortonworks Sandbox wi
root@sandbox-hdp:/usr/
root@sandbox-hdp:/usr/
13.82.28.75:4200
Most Visited
Getting Started
NYU Email
> django.contrib.contenttypes
> django.contrib.sessions
> django.contrib.sites
> django.contrib.admin
> django_extensions
> south
> about
> filebrowser
> help
> proxy
Migrated:
- desktop
- djcelery
- beeswax
- jobsub
- oozie
- pig
- useradmin
No fixtures found.
Validating models...
0 errors found
Django version 1.2.3, using settings 'desktop.settings'
Development server is running at http://127.0.0.1:8000/
Quit the server with CONTROL-C.
^C
Server stopped.
Note that the test database, '/usr/lib/hue/desktop/desktop-test.db', has not been deleted. Y
[root@sandbox-hdp hue]# cp /usr/lib/hue/desktop/desktop-test.db /var/lib/hue/desktop.db
cp: overwrite `/var/lib/hue/desktop.db'? y
[root@sandbox-hdp hue]# cp /etc/hue/conf.empty/hue.ini /etc/hue/conf.empty/hue.ini.bak
[root@sandbox-hdp hue]# vi /etc/hue/conf.empty/hue.ini
[root@sandbox-hdp hue]#

Start Hue


Inbound security ru
Hortonworks Sandl
root@sandbox-hdp
root@sandbox-hdp
New Tab
13.82.28.75:4200
Search
Most Visited
Getting Started
NYU Email
0 errors found
Django version 1.2.3, using settings 'desktop.settings'
Development server is running at http://127.0.0.1:8000/
Quit the server with CONTROL-C.
^C
Server stopped.
Note that the test database, '/usr/lib/hue/desktop/desktop-test.db', has not been deleted. You
[root@sandbox-hdp hue]# cp /usr/lib/hue/desktop/desktop-test.db /var/lib/hue/desktop.db
cp: overwrite `/var/lib/hue/desktop.db'? y
[root@sandbox-hdp hue]# cp /etc/hue/conf.empty/hue.ini /etc/hue/conf.empty/hue.ini.bak
[root@sandbox-hdp hue]# vi /etc/hue/conf.empty/hue.ini
[root@sandbox-hdp hue]# service hue start
Detecting versions of components...
HUE_VERSION=2.6.1-91
Starting hue: [OK]
[root@sandbox-hdp hue]#

Username: hue, Password: 1111



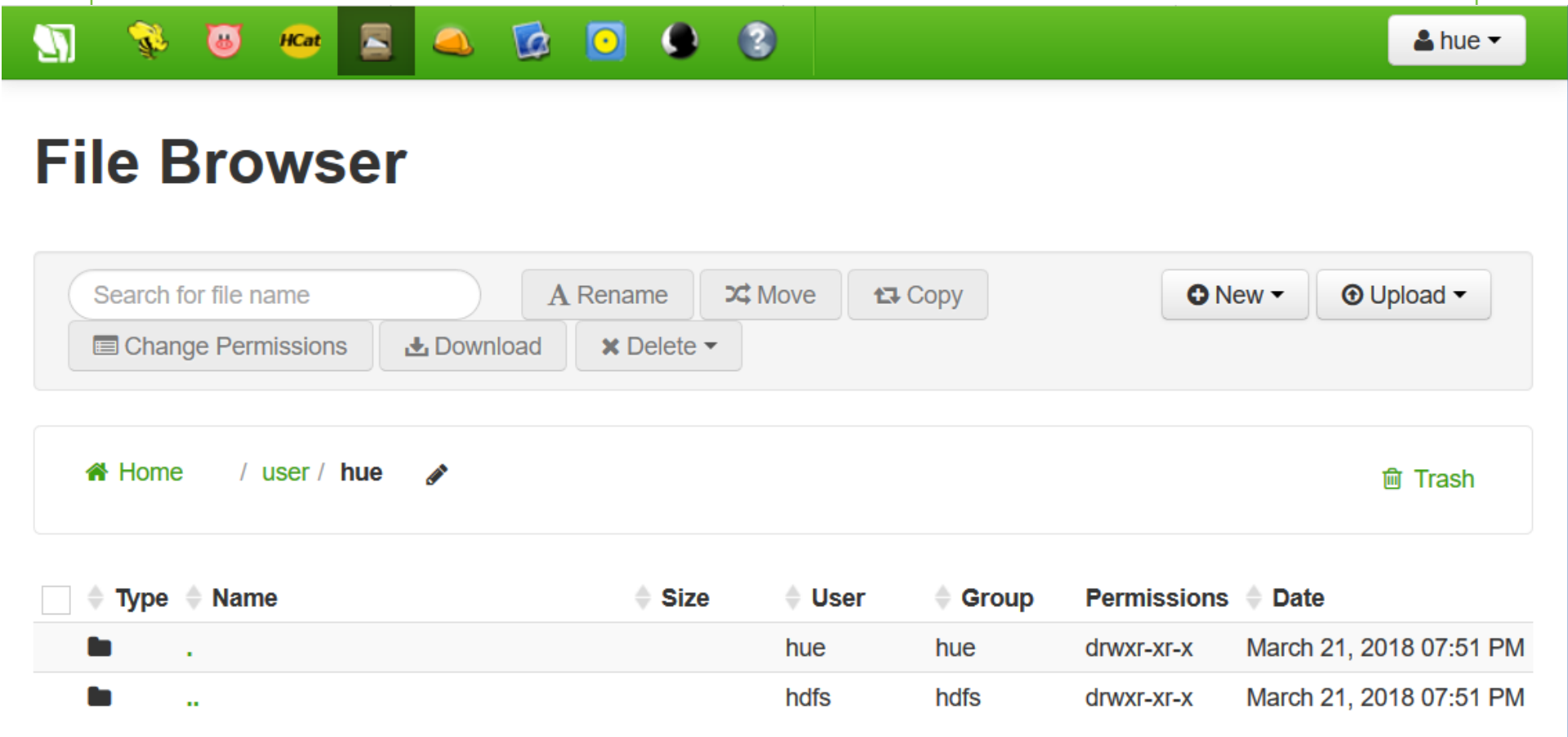


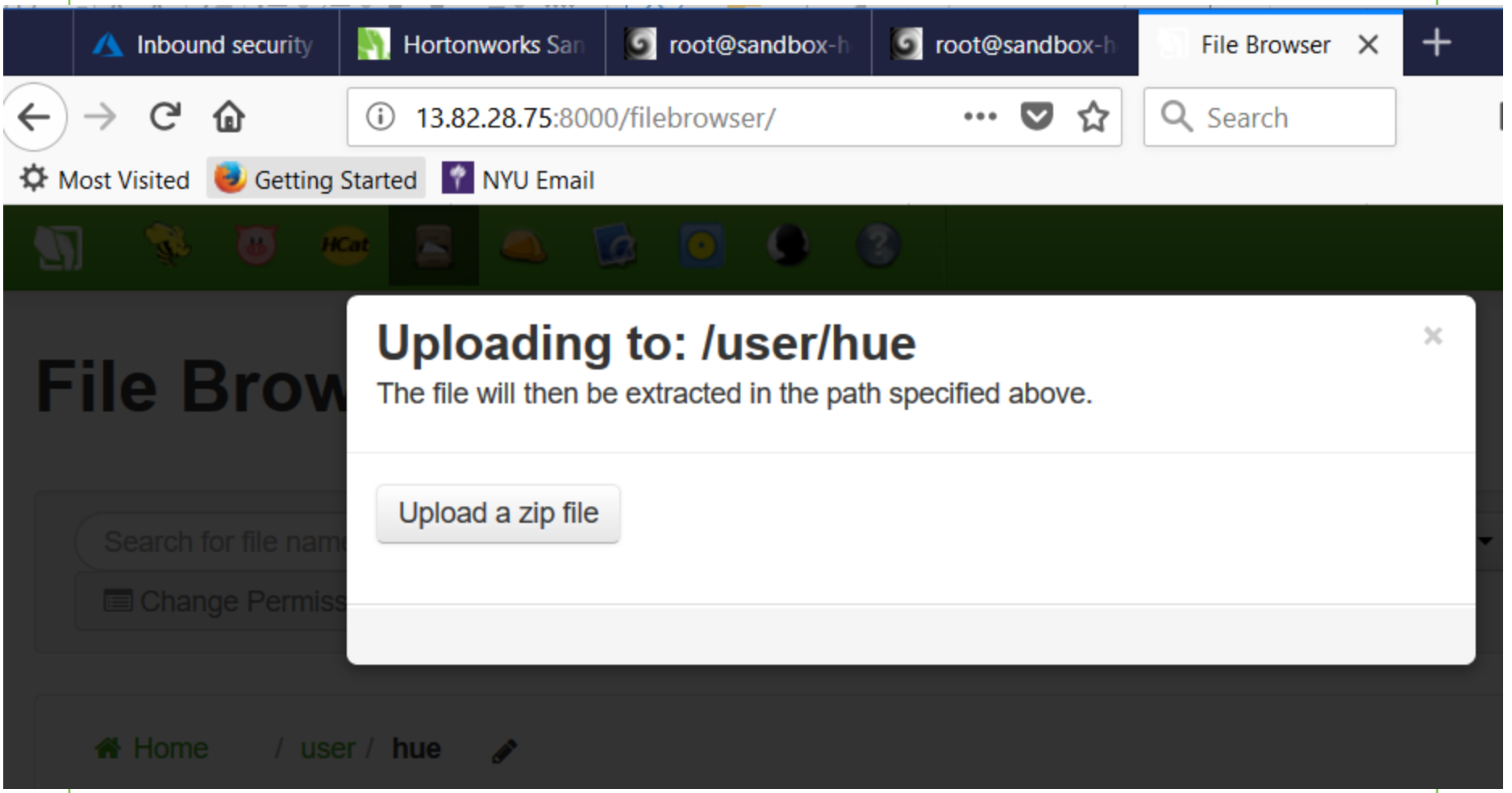
Hue

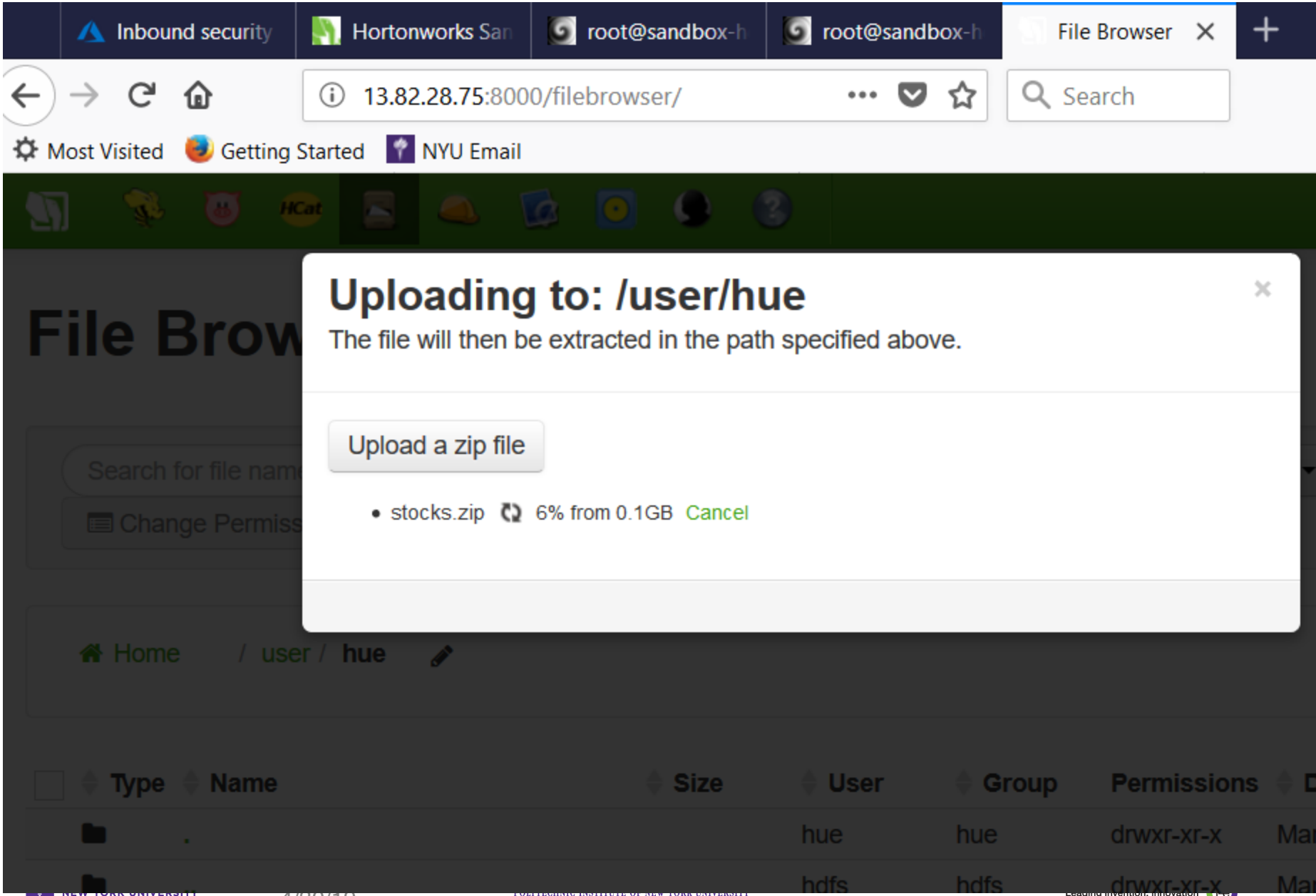
Component	Version
Tutorials	2.0.005
Hue	2.6.1-91
Sandbox Build	f1dc3df 09:23 03-04-15

Load Raw data into Hadoop File System

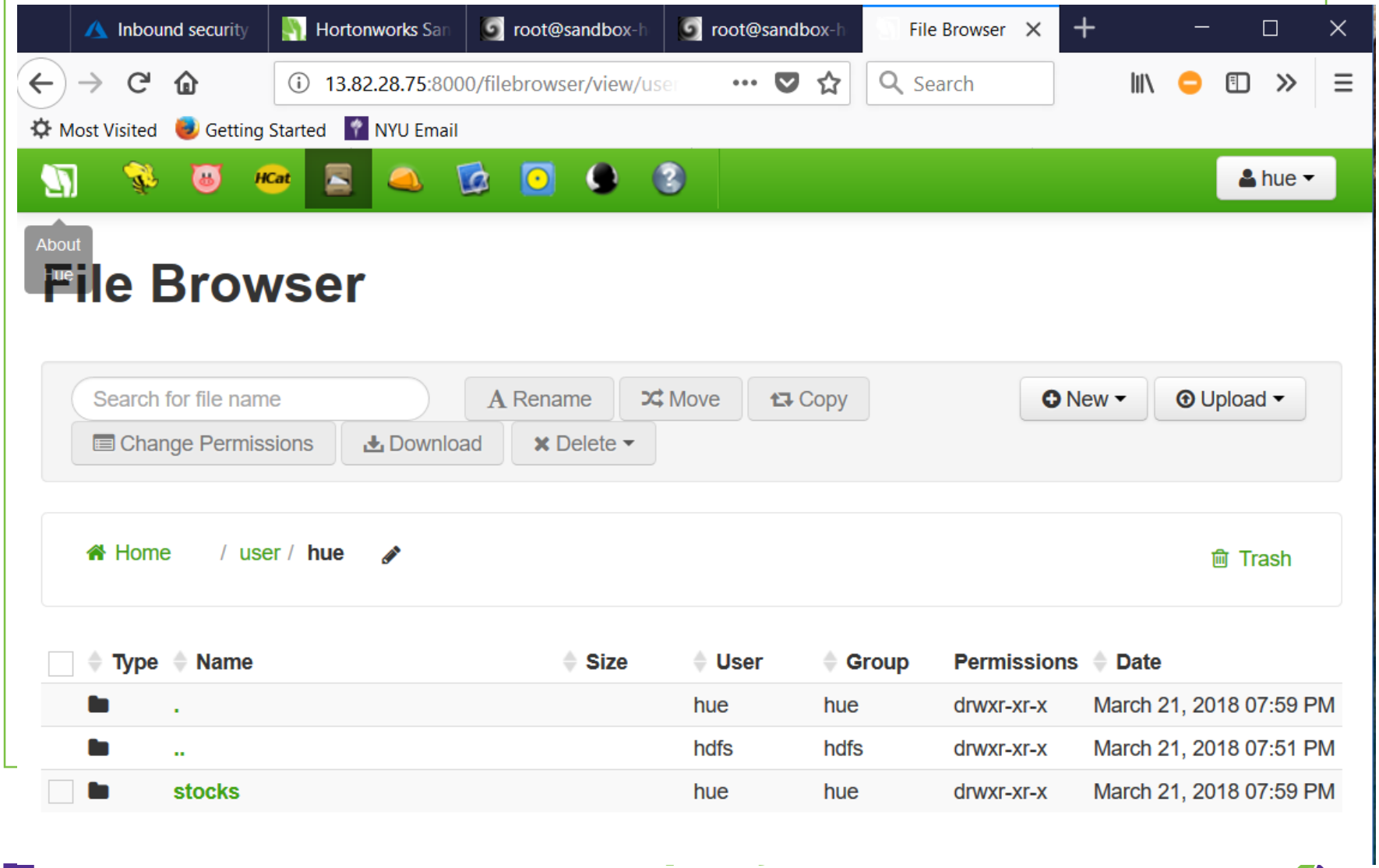
- Download the stocks.zip file from our class Web site.
- Click File Brower on your HDP Portal, and Upload the zip file into your HDP File System.




Inbound security
Hortonworks San
root@sandbox-h
root@sandbox-h
File Browser
13.82.28.75:8000/filebrowser/
Search
Most Visited
Getting Started
NYU Email
HCat
File Brow
Uploading to: /user/hue
The file will then be extracted in the path specified above.
Upload a zip file
Search for file nam
Change Permiss
Home / user / hue

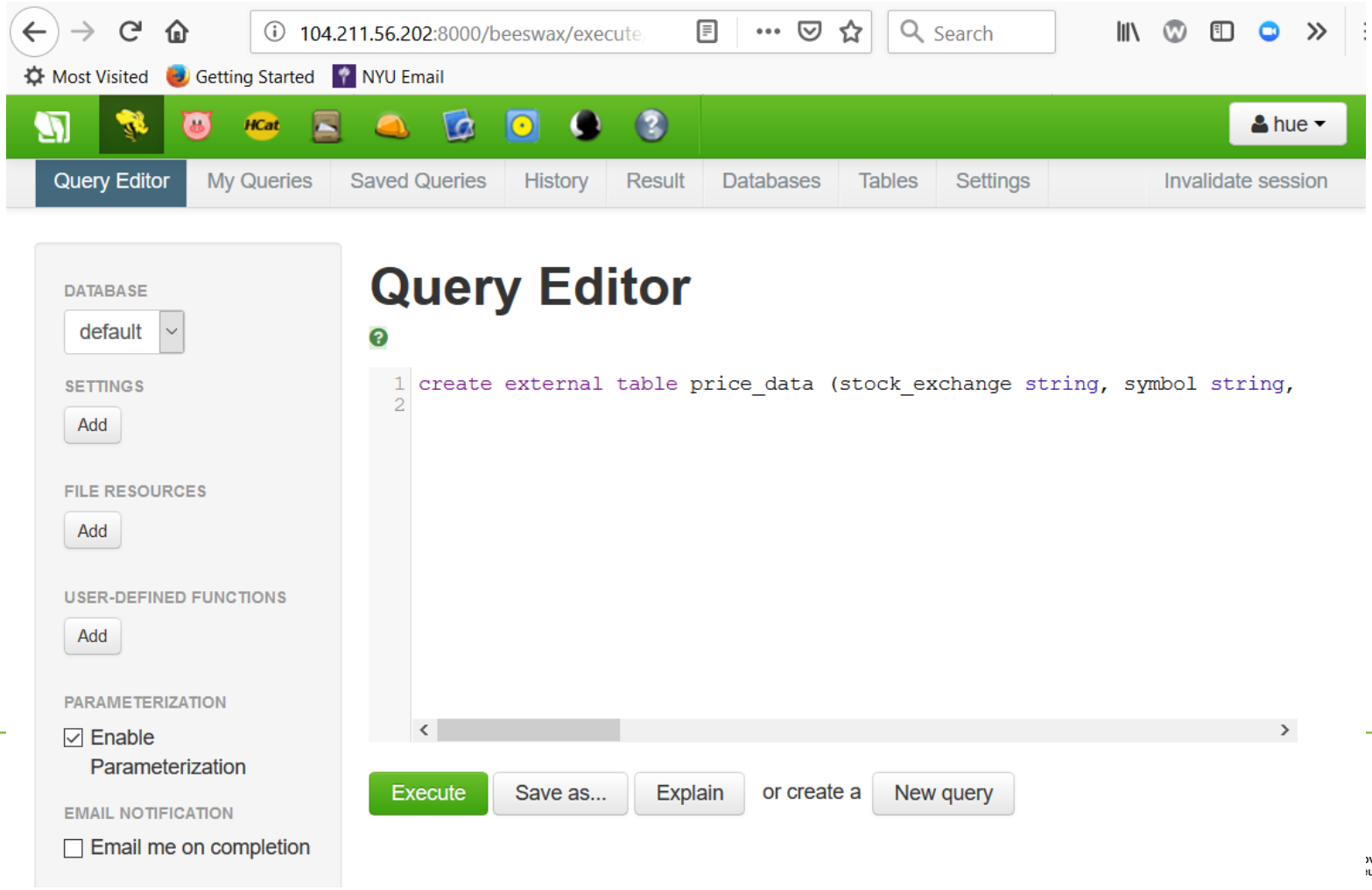

Inbound security
Hortonworks San
root@sandbox-h
root@sandbox-h
File Browser
13.82.28.75:8000/filebrowser/
Search
Most Visited
Getting Started
NYU Email
HCat
File Brow
Uploading to: /user/hue
The file will then be extracted in the path specified above.
Upload a zip file
stocks.zip 6% from 0.1GB Cancel
Search for file nam
Change Permiss
Home / user / hue
Type Name Size User Group Permissions
hue hue drwxr-xr-x
hdfs hdfs drwxr-xr-x

File Loading is complete



Create tables in the default database

- In Query Editor, type DDL to create table price_data, dividends_data and yearly_aggregates shown on the next 2 slides.

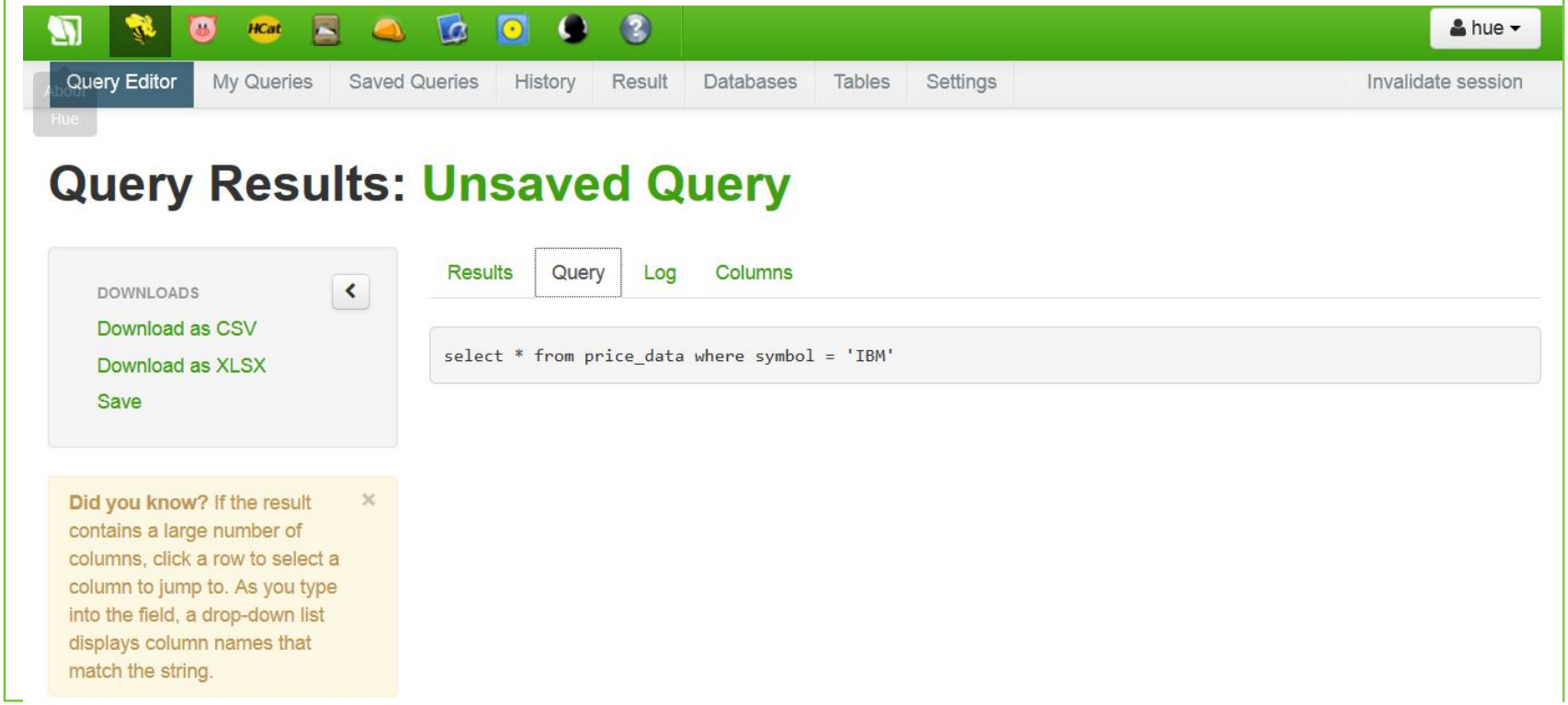


The DDL to create tables price_data and dividends_data

- create external table price_data (stock_exchange string, symbol string, trade_date string, open float, high float, low float, close float, volume int, adj_close float) row format delimited fields terminated by ',' stored as textfile location '/user/hue/stocks/stocks/prices';
- create external table dividends_data (stock_exchange string, symbol string, trade_date string, dividend float) row format delimited fields terminated by ',' stored as textfile location '/user/hue/stocks/stocks/dividends';

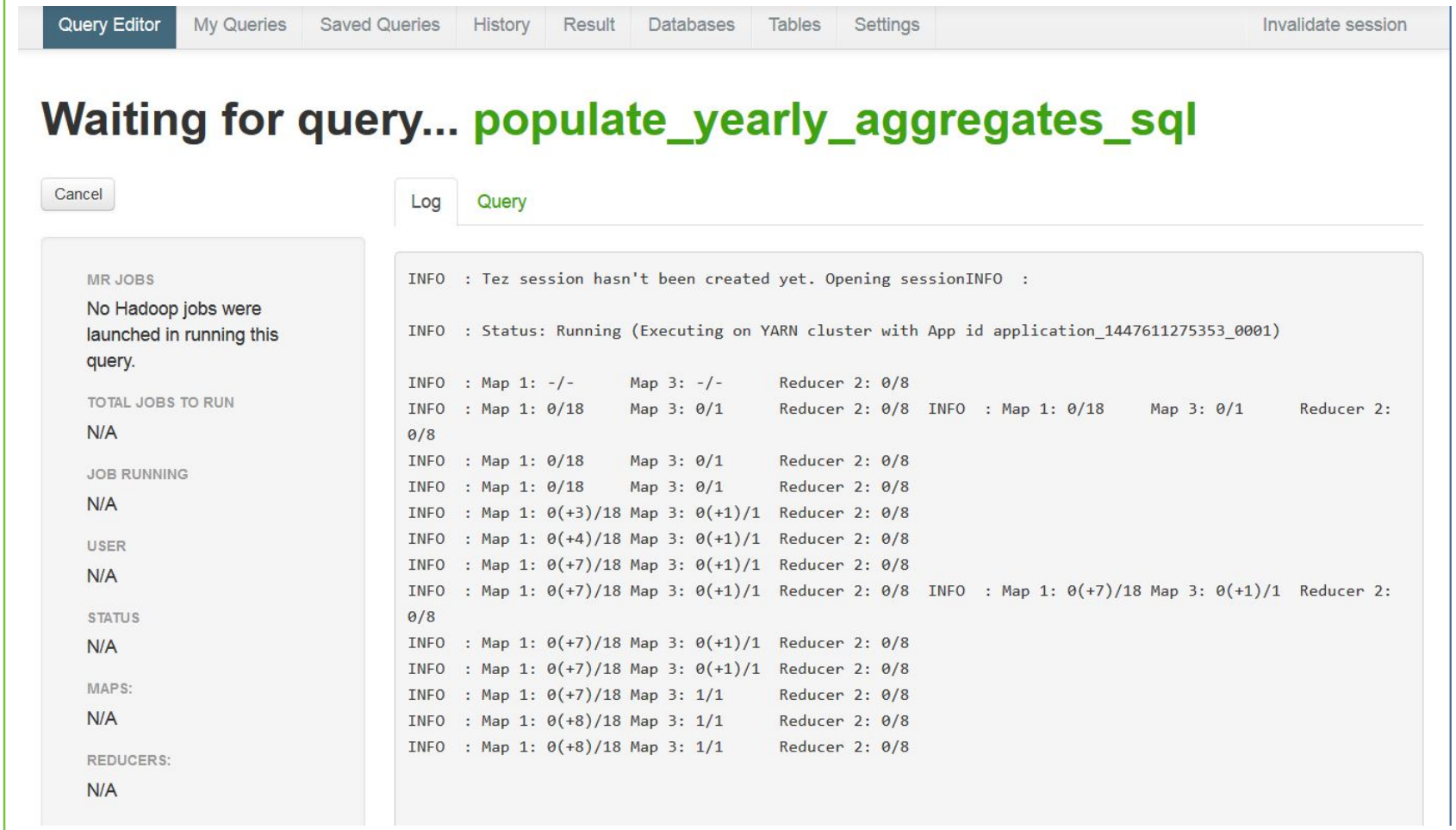
Verify your tables price_data and dividends_data

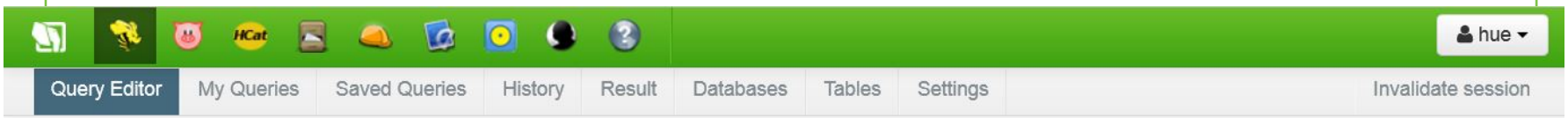
- select * from price_data where symbol = 'IBM'
- select * from dividends_data where symbol = 'IBM'



Create and Populate Table yearly_aggregates

- Create the table yearly_aggregates:
  - create table yearly_aggregates (symbol string, year string, high float, low float, average_close float, total_dividends float) row format delimited fields terminated by ',' stored as textfile location '/user/hue/stocks/stocks/stock_aggregates';
- Populate the table:
  - insert overwrite table yearly_aggregates select a.symbol, year(a.trade_date), max(a.high), min(a.low), avg(a.close), sum(b.dividend) from price_data a left outer join dividends_data b on (a.symbol = b.symbol and a.trade_date = b.trade_date) group by a.symbol, year(a.trade_date);
- Verify the table:
  - select * from yearly_aggregates where symbol = 'IBM'


Query Editor
My Queries
Saved Queries
History
Result
Databases
Tables
Settings
Invalidate session
Waiting for query... populate_yearly_aggregates_sql
Cancel
Log
Query
MR JOBS
No Hadoop jobs were launched in running this query.
TOTAL JOBS TO RUN
N/A
JOB RUNNING
N/A
USER
N/A
STATUS
N/A
MAPS:
N/A
REDUCERS:
N/A
INFO : Tez session hasn't been created yet. Opening sessionINFO :
INFO : Status: Running (Executing on YARN cluster with App id application_1447611275353_0001)
INFO : Map 1: -/- Map 3: -/- Reducer 2: 0/8
INFO : Map 1: 0/18 Map 3: 0/1 Reducer 2: 0/8 INFO : Map 1: 0/18 Map 3: 0/1 Reducer 2: 0/8
INFO : Map 1: 0/18 Map 3: 0/1 Reducer 2: 0/8
INFO : Map 1: 0/18 Map 3: 0/1 Reducer 2: 0/8
INFO : Map 1: 0(+3)/18 Map 3: 0(+1)/1 Reducer 2: 0/8
INFO : Map 1: 0(+4)/18 Map 3: 0(+1)/1 Reducer 2: 0/8
INFO : Map 1: 0(+7)/18 Map 3: 0(+1)/1 Reducer 2: 0/8
INFO : Map 1: 0(+7)/18 Map 3: 0(+1)/1 Reducer 2: 0/8 INFO : Map 1: 0(+7)/18 Map 3: 0(+1)/1 Reducer 2: 0/8
INFO : Map 1: 0(+7)/18 Map 3: 0(+1)/1 Reducer 2: 0/8
INFO : Map 1: 0(+7)/18 Map 3: 0(+1)/1 Reducer 2: 0/8
INFO : Map 1: 0(+7)/18 Map 3: 1/1 Reducer 2: 0/8
INFO : Map 1: 0(+8)/18 Map 3: 1/1 Reducer 2: 0/8
INFO : Map 1: 0(+8)/18 Map 3: 1/1 Reducer 2: 0/8



Query Results: Unsaved Query

DOWNLOADS

Download as CSV

Download as XLSX

Save

Did you know? If the result contains a large number of columns, click a row to select a column to jump to. As you type into the field, a drop-down list displays column names that match the string.

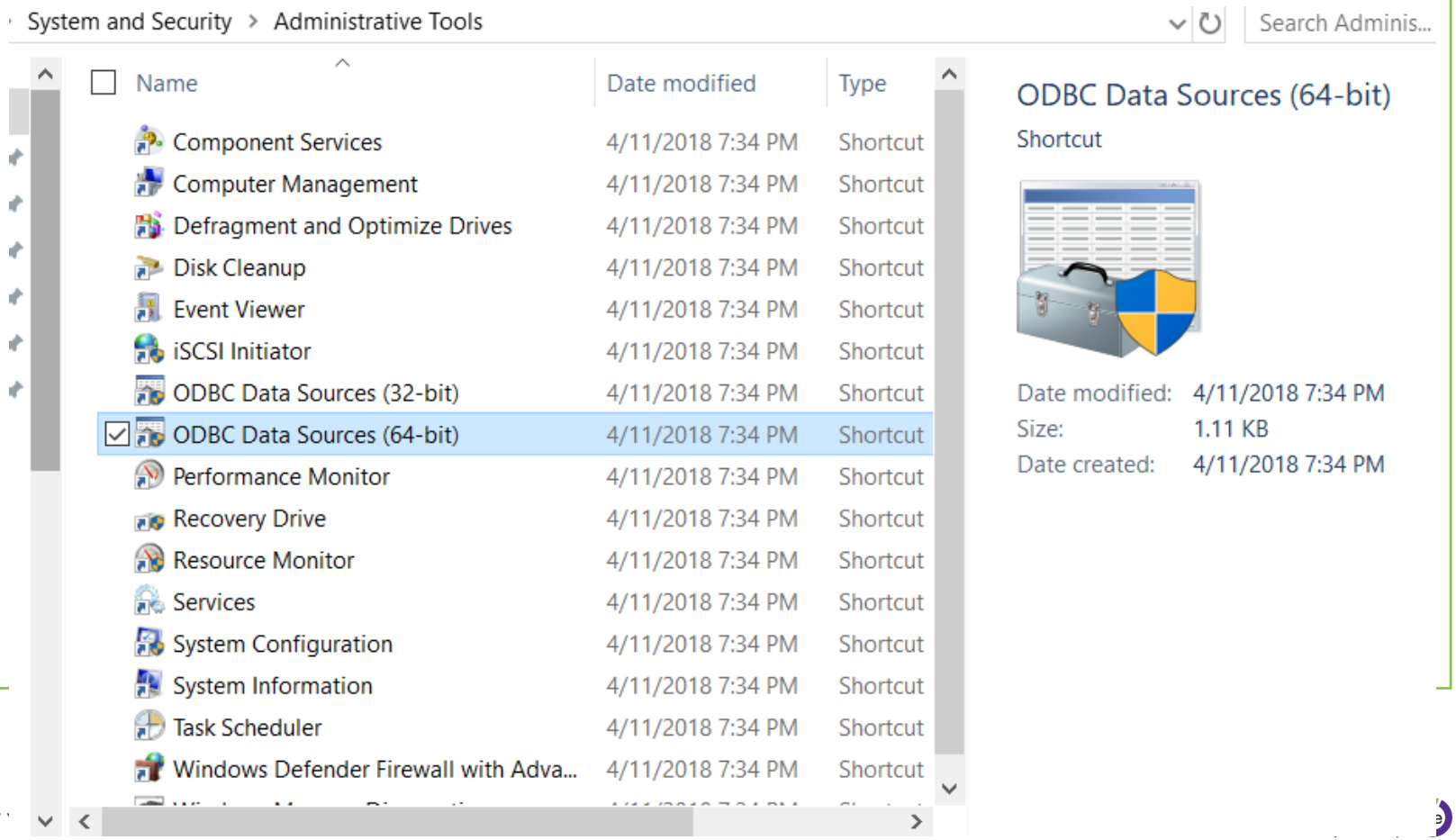
Results | Query | Log | Columns

	yearly_aggregates.symbol	yearly_aggregates.year	yearly_aggregates.high	yearly_aggregates.low	y
14	IBM	1962	578.5	300.0	4
15	IBM	1963	509.5	384.25	4
37	IBM	1964	601.5	407.0	4
19	IBM	1965	554.0	404.0	4
27	IBM	1966	562.0	289.5	4
9	IBM	1967	648.0	362.5	5
28	IBM	1968	649.880004883	304.5	4
29	IBM	1969	368.75	291.75	3
30	IBM	1970	387.0	218.75	2
31	IBM	1971	365.75	283.25	3
38	IBM	1972	426.75	331.75	3
32	IBM	1973	456.5	235.130004883	3
33	IBM	1974	254.0	150.5	2

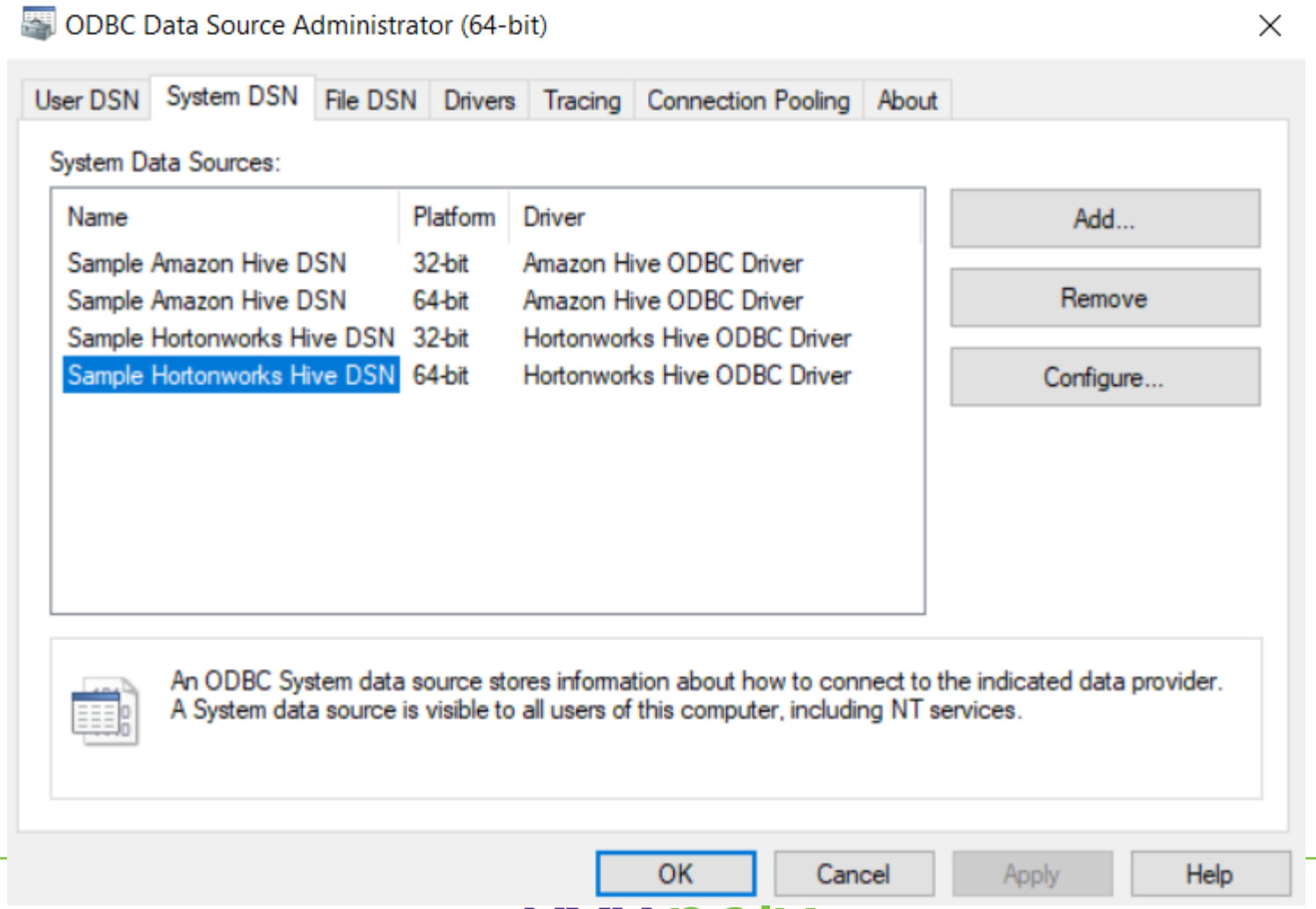
- For Apple Mac Laptops, go to page 57, Instructions of hortonworks Hive ODBC driver installation, and rest procedures.

Connect HDP VM with a Python program via ODBC

- For Windows laptop, download the ODBC Driver, **HortonworksHiveODBC64**, from our class Web site. Click it to install.
- Then go to ODBC Driver Manager to configure your ODBC connection.

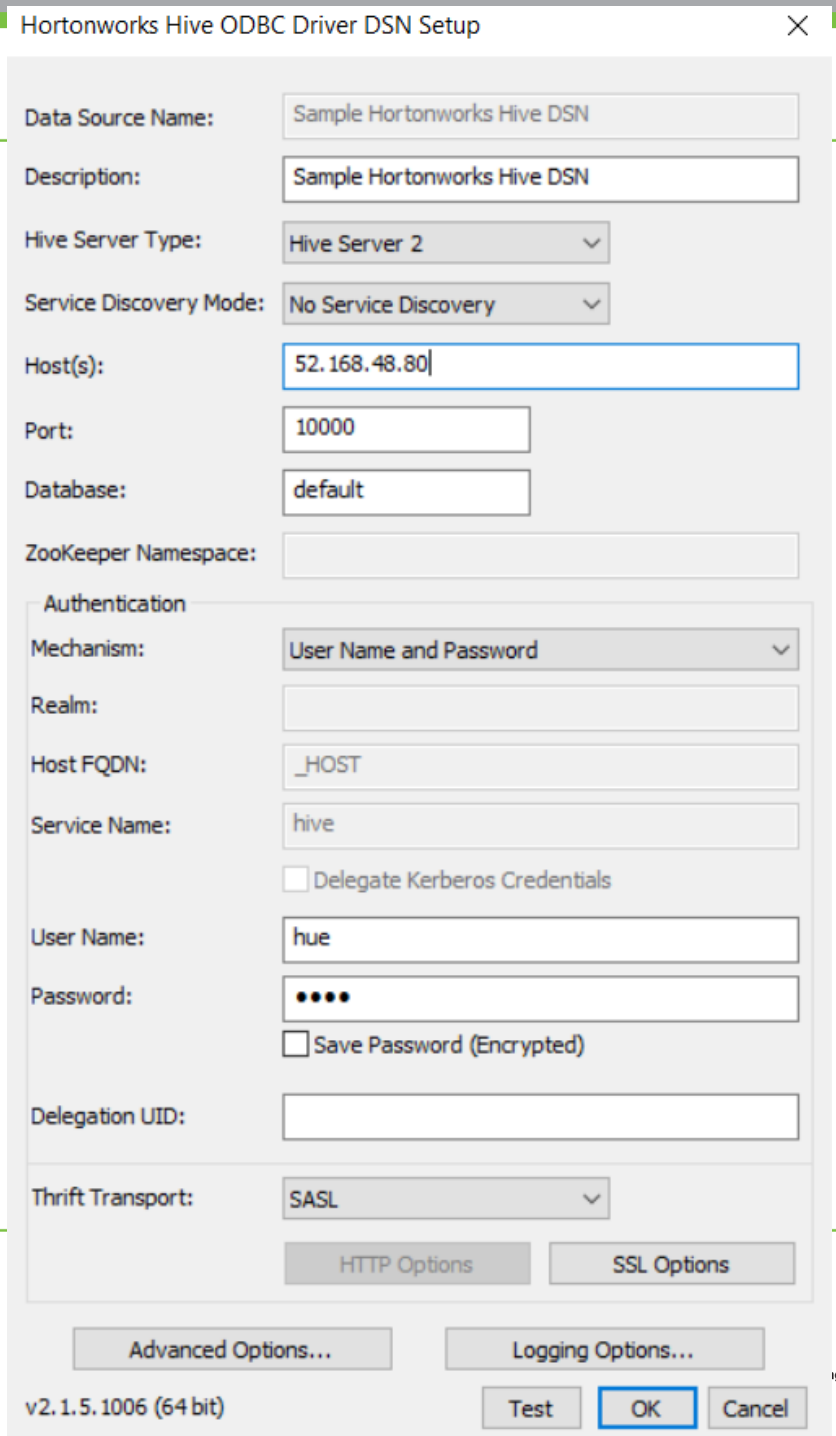


Configure Hortonworks HIVE ODBC Driver

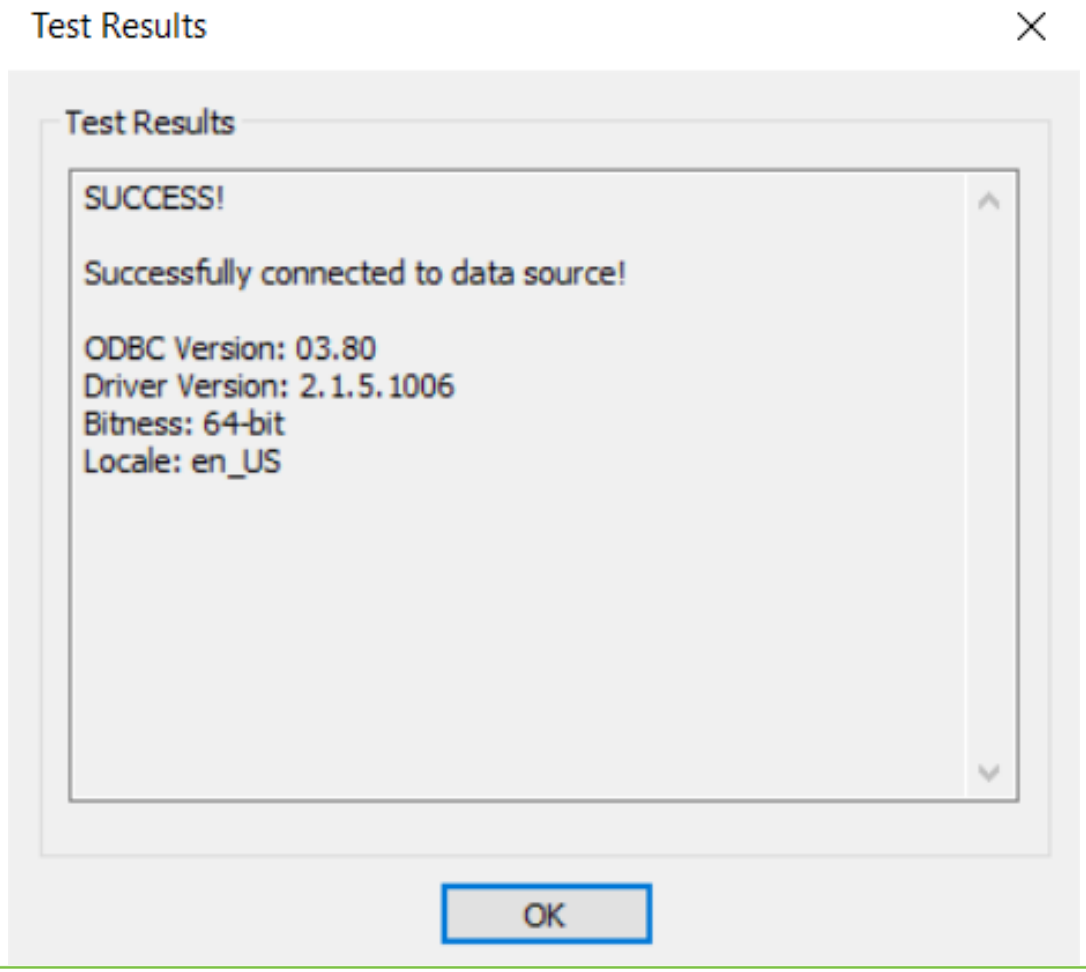


Configure Hortonworks HIVE ODBC Driver (Continue)

- Use the IP Address for your VM
- Use your HUE password 1111
- Then test connection to your Azure HDP VM.



Hortonworks Hive ODBC Driver Data Source Test



Pull Data from Azure Cloud to your Python Program

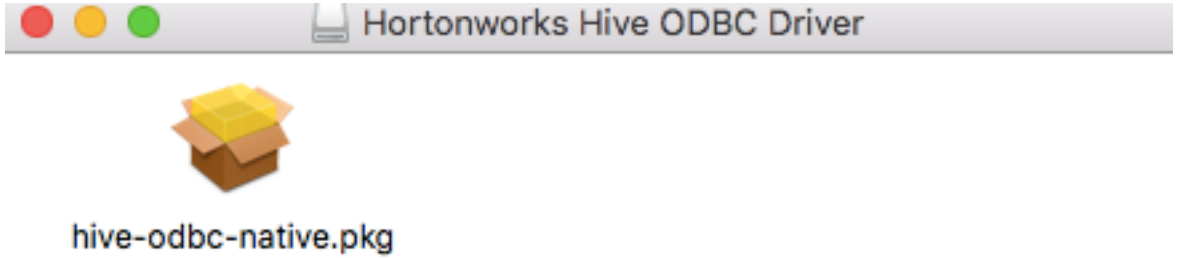
- Create a python program to pull selected data from your dividend table in Cloud

```
import pyodbc

pyodbc.autocommit = True
conn = pyodbc.connect("DSN=Sample Hortonworks Hive DSN;", autocommit=True)

cursor = conn.cursor();

cursor.execute("select symbol,trade_date as dividend_date, dividend from
default.dividends_data where symbol = 'IBM' and year(trade_date) = '1990'")

result = cursor.fetchall()
for r in result:
    print(r)
```

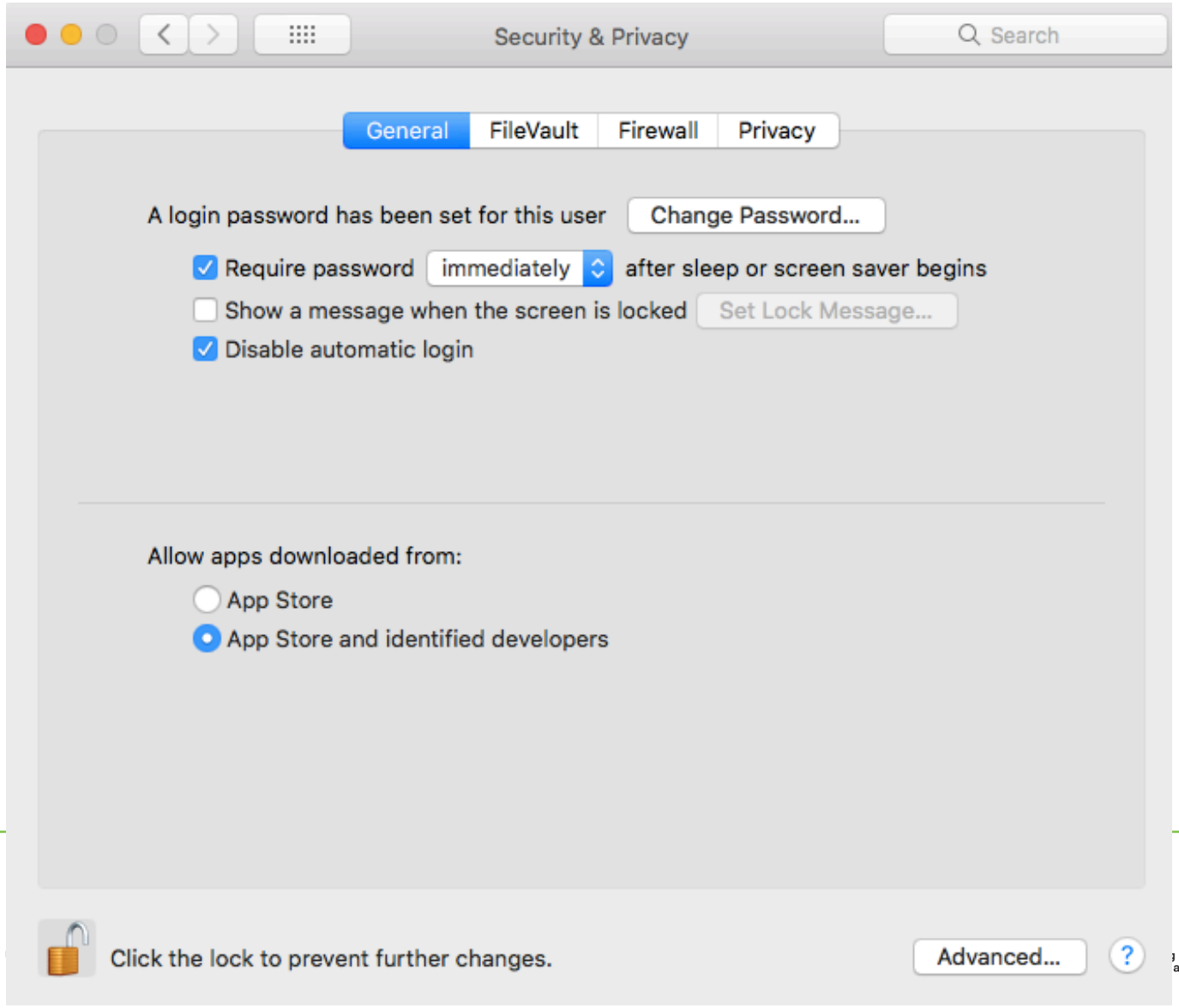
```
('IBM', '1990-11-05', 0.30250000953674316)
('IBM', '1990-08-06', 0.30250000953674316)
('IBM', '1990-05-04', 0.30250000953674316)
('IBM', '1990-02-05', 0.30250000953674316)
```

- Page 58 – Page 62 are for Apple MAC OS X.
- Students with Windows PC ignore these slides.

Step1: Download the Hortonworks Hive ODBC Driver from our class Web site



Allow program not from App Store to be installed



Step 2: Download ODBC configuration file

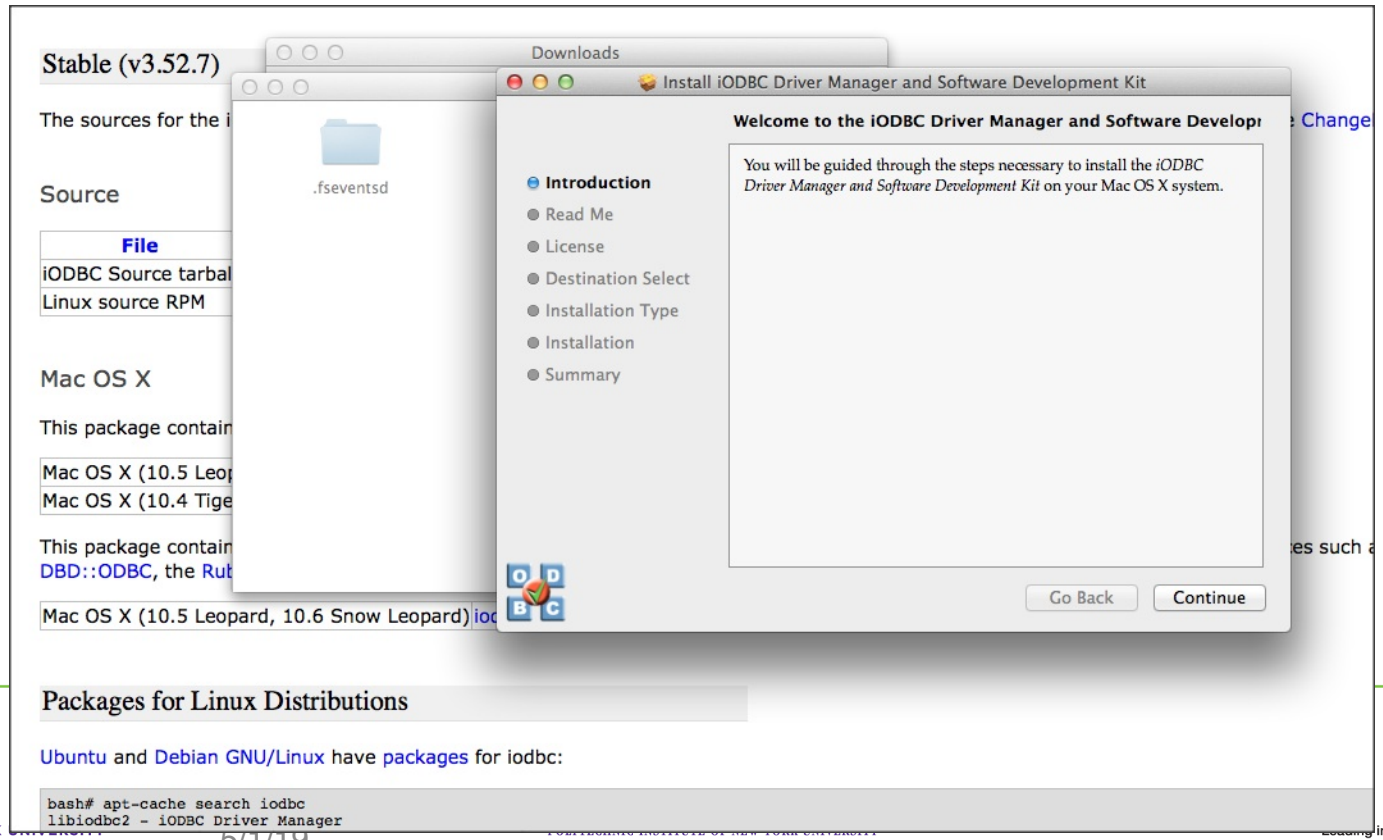
- From ODBC Driver for MAC folder of our class Web site, download the file: odbc.ini
- Run the following commands via a terminal:
  - $ cd ~
  - $ cd Library/ODBC
  - $ sudo mv odbc.ini odbc.ini.bak
  - $ cd ~
  - $ cd Downloads/
  - $ sudo cp odbc.ini ../Library/ODBC/
- Use vi to change the host to the IP address of your VM

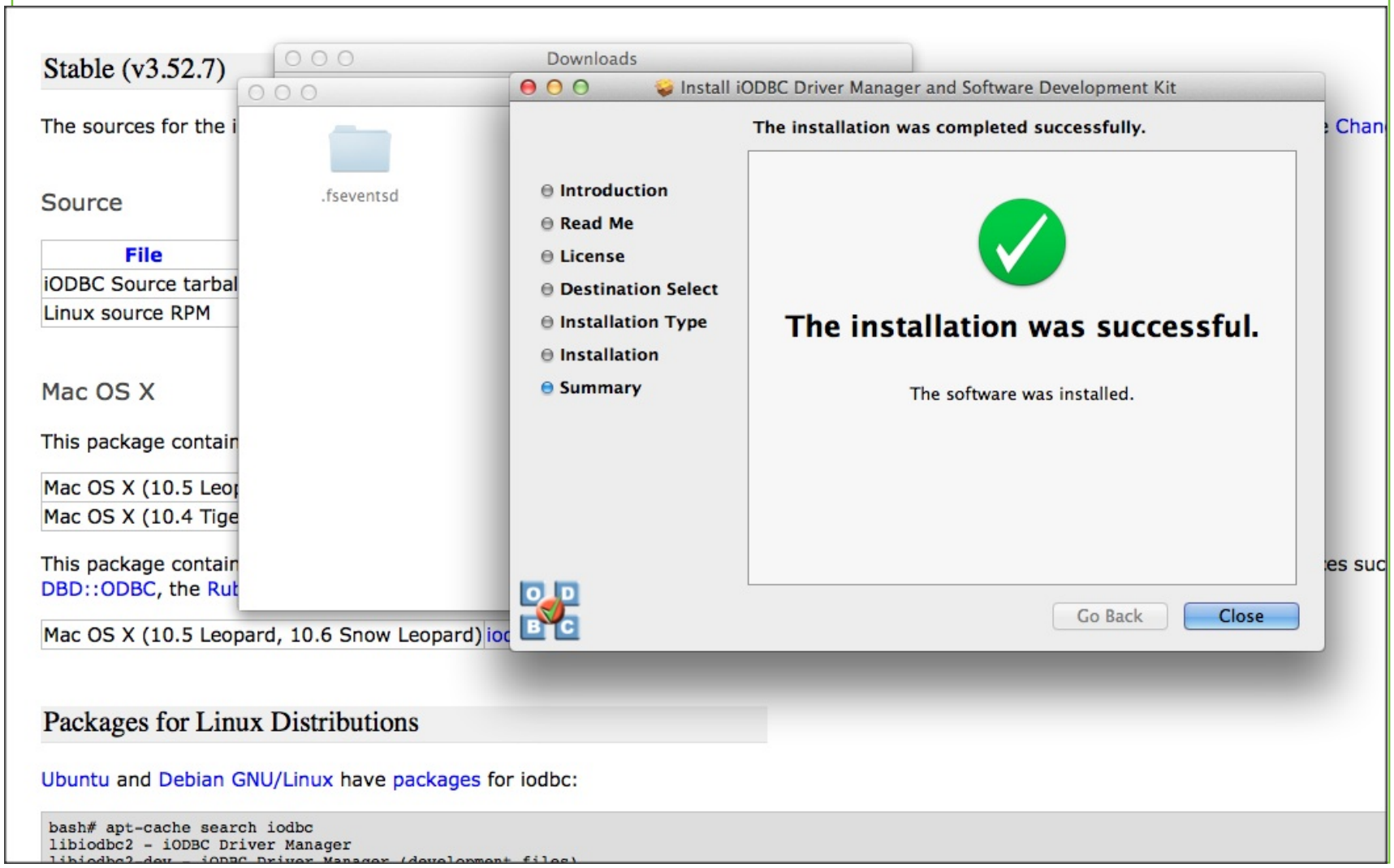
Step 3 Verify your Anaconda, Python Version and pyodbc

- $ echo $PATH
  - Make sure you have anaconda3/bin in your PATH
- $ python -V
  - Python 3.6.4 :: Anaconda custom (64-bit) or higher
- $ Check your Anaconda env has the package pyodbc

Step 4: Download and Install the iODBC Driver Manager for Mac OS X

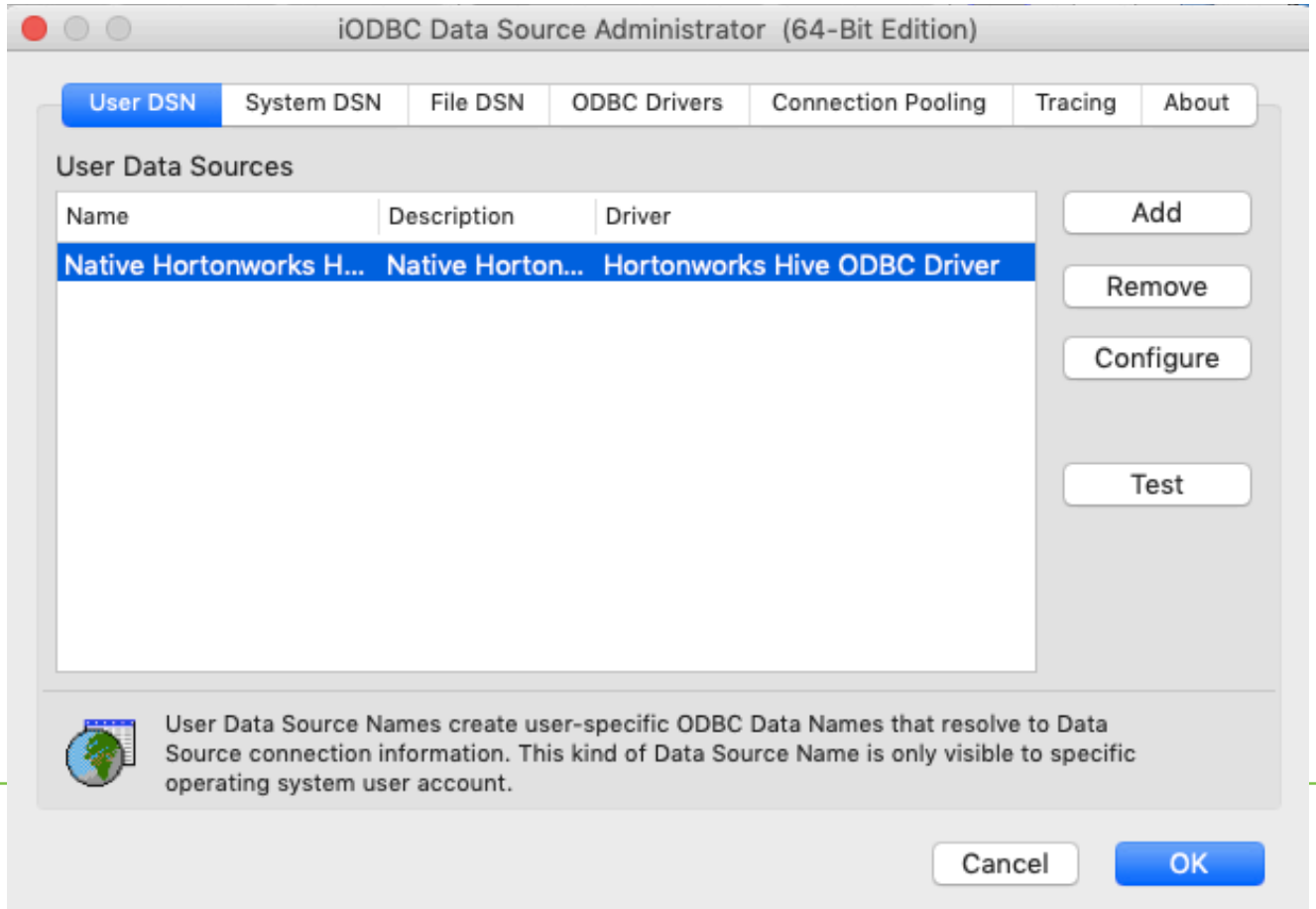
- Download the iODBC Driver Manager, **mxkozzzz.dmg**, from our class Web site for installation.




Stable (v3.52.7)
The sources for the i
Source
File
iODBC Source tarbal
Linux source RPM
Mac OS X
This package contain
Mac OS X (10.5 Leop
Mac OS X (10.4 Tige
This package contain
DBD::ODBC, the Rut
Mac OS X (10.5 Leopard, 10.6 Snow Leopard)
Packages for Linux Distributions
Ubuntu and Debian GNU/Linux have packages for iodbc:
bash# apt-cache search iodbc
libiodbc2 - iODBC Driver Manager
Downloads
.fseventsd
Install iODBC Driver Manager and Software Development Kit
The installation was completed successfully.
Introduction
Read Me
License
Destination Select
Installation Type
Installation
Summary
The installation was successful.
The software was installed.
Go Back
Close

Test the connection to your Azure Hortonswork

- Start iODBC Administrator64, and click Test. You should see message saying connection is successful.



Pull Data from Azure Cloud to your Python Program

- Create a python program to pull selected data from your dividend table in Cloud

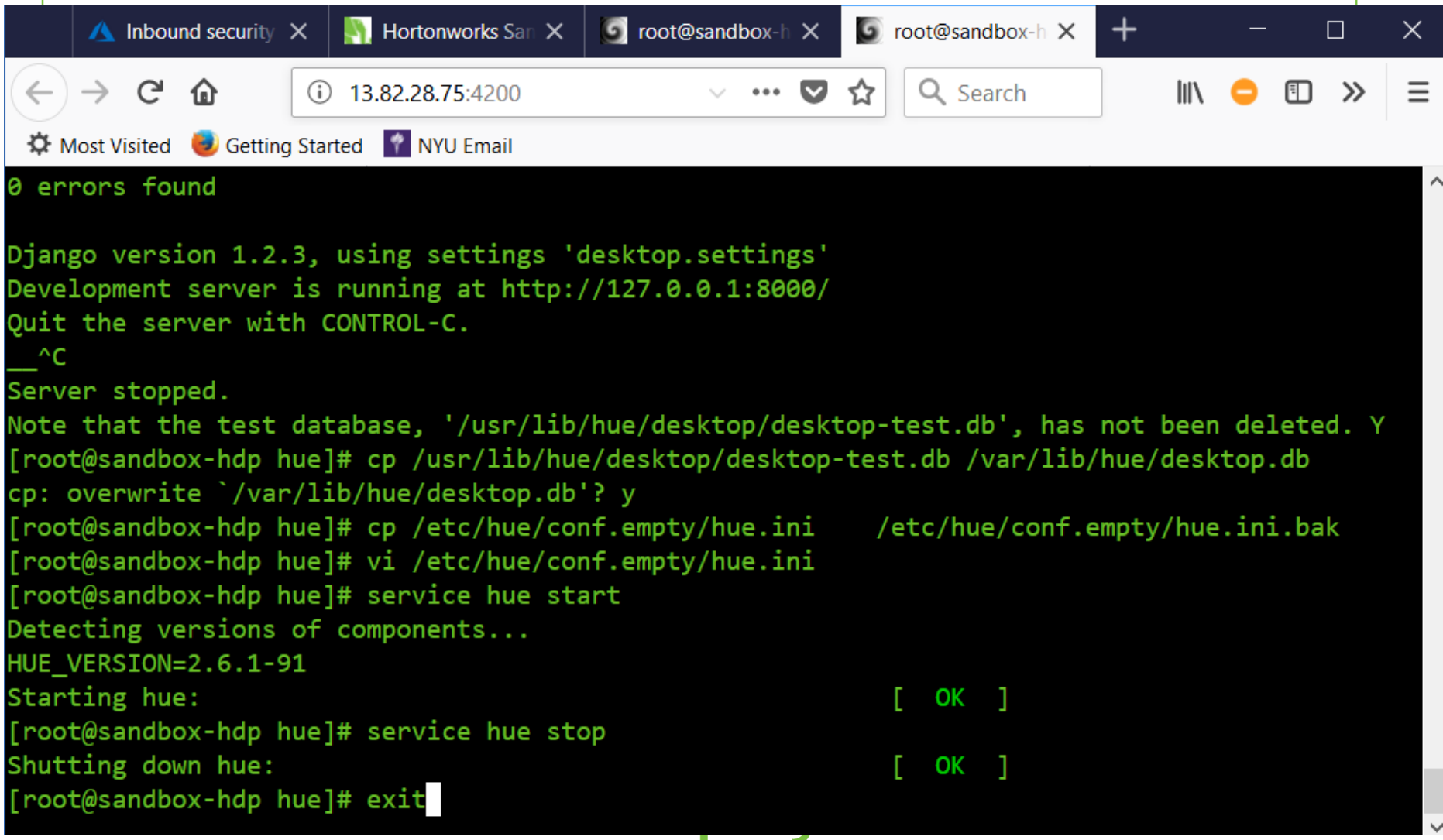
```
import pyodbc

pyodbc.autocommit = True
conn = pyodbc.connect("DSN=Native Hortonworks Hive DSN;", autocommit=True)

cursor = conn.cursor();

cursor.execute("select symbol,trade_date as dividend_date, dividend from
default.dividends_data where symbol = 'IBM' and year(trade_date) = '1990'")

result = cursor.fetchall()
for r in result:
    print(r)
```

```
('IBM', '1990-11-05', 0.30250000953674316)
('IBM', '1990-08-06', 0.30250000953674316)
('IBM', '1990-05-04', 0.30250000953674316)
('IBM', '1990-02-05', 0.30250000953674316)
```

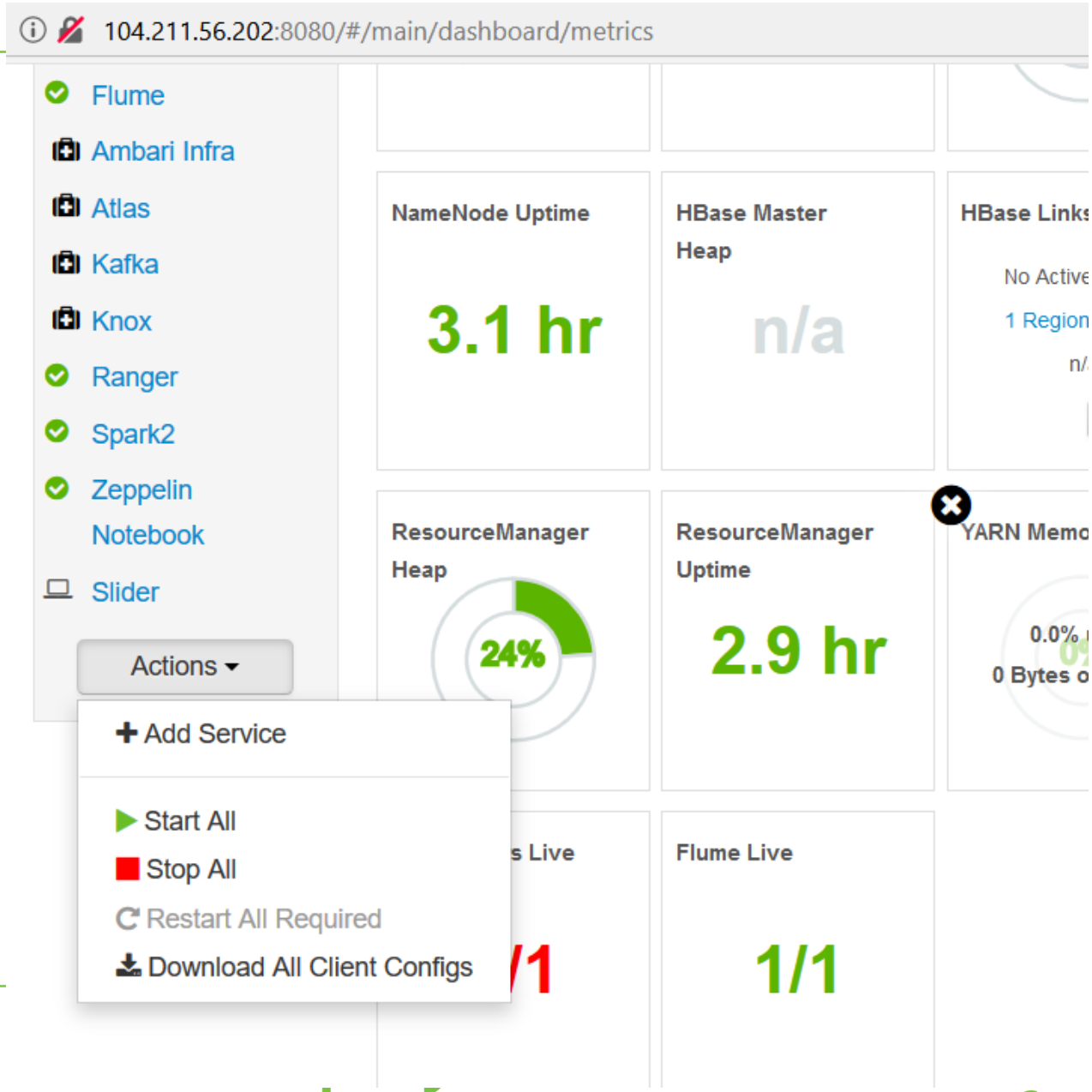
Homework Assignment 6

- Connect your Python program to your Hortonworks database in Azure to pull IBM yearly dividend from yearly_aggregates table for the year assigned to you.
-
- Please go to resources section of our class Web to see the year assigned to you in the file Dividend_Year_for_IBM.
-
- Submit your Python codes with the results pasted under the codes such as
-
- '''
-
- ('1977', 'IBM', Decimal('0.625'))
-
- '''

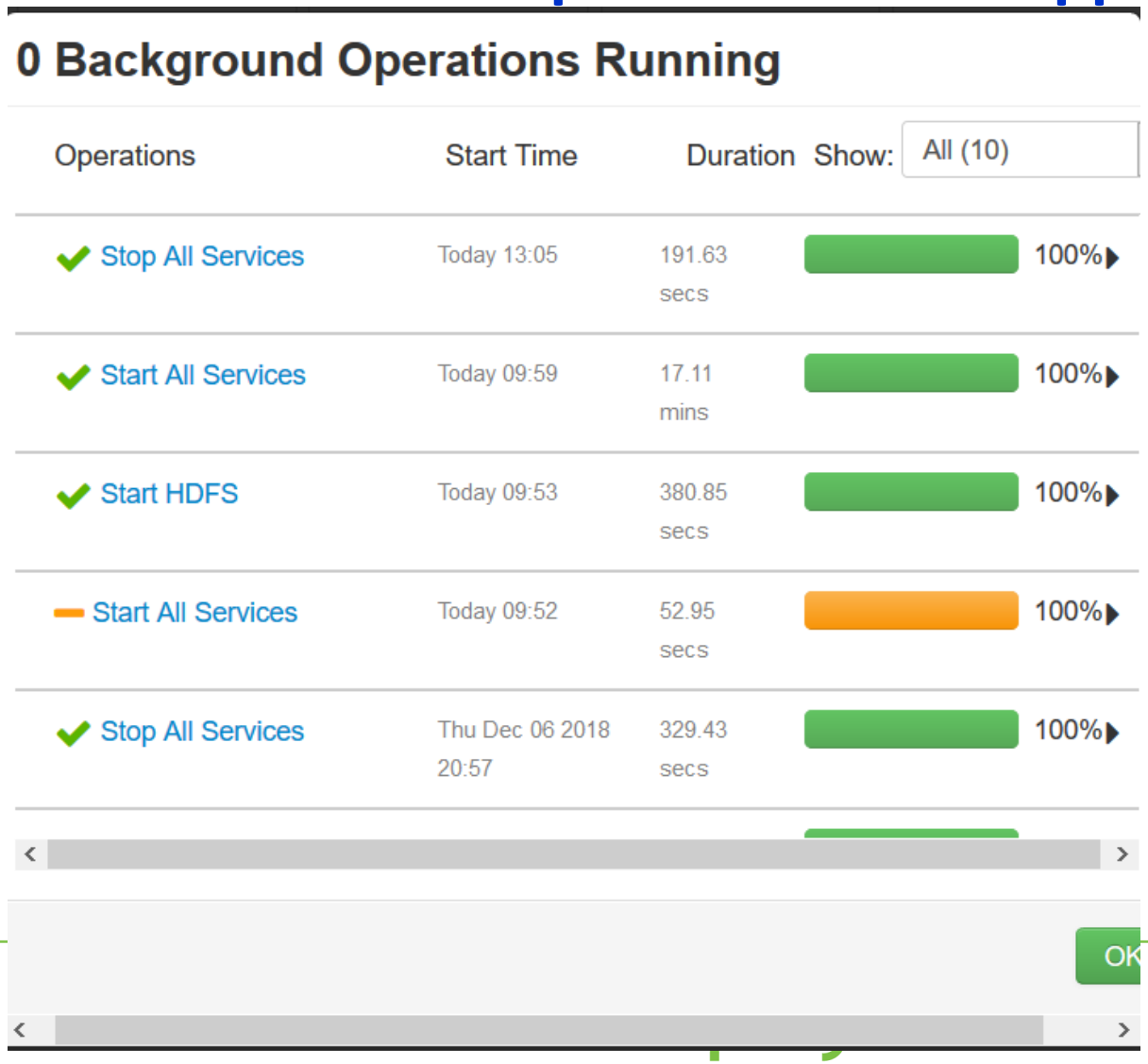
Stop Hue Service and Exit



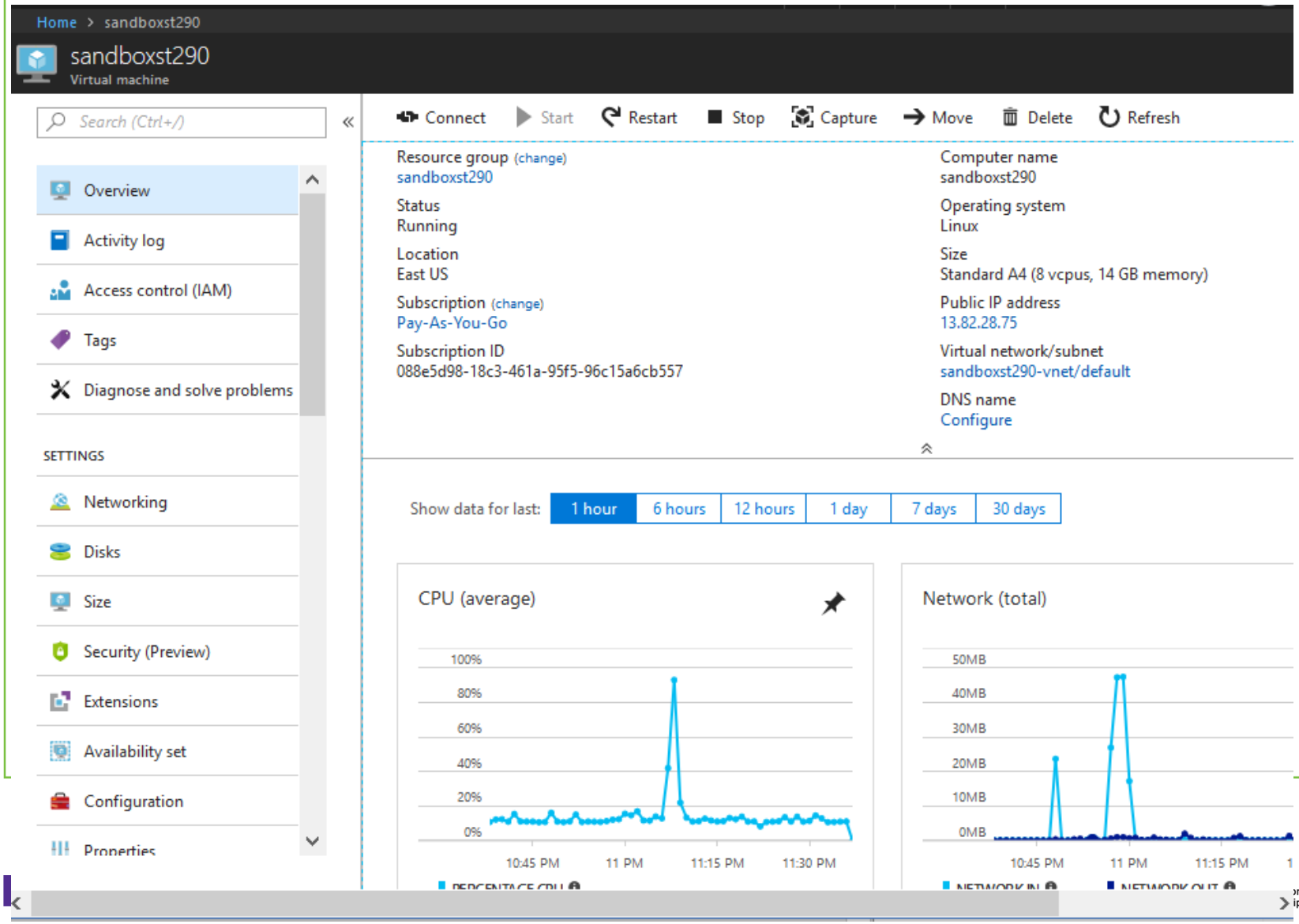
Stop all Hortonworks Processes by Click Stop ALL



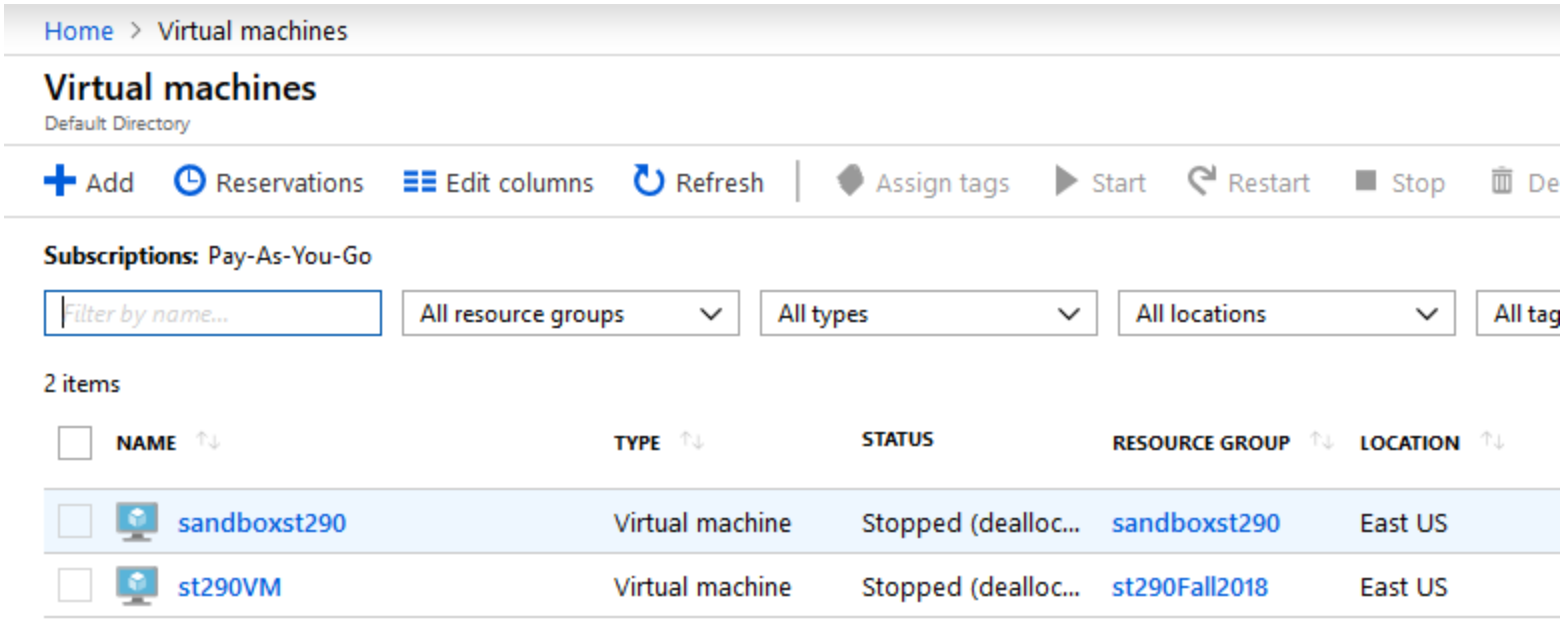
All the Hortonworks processes are stopped



Stop your VM by clicking Stop!



Verify your VM is stopped



References:

- Getting Started with Hortonworks Sandbox on Azure. http://hortonworks.com/blog/hortonworks-sandbox-azure/
- Deriving Business Insight from Data using Microsoft Excel and Hortonworks Data Platform.
- http://hortonworks.com/hadoop-tutorial/partner-tutorial-microsoft/
- How to Install and Configure the Hortonworks ODBC driver on Mac OS X. http://hortonworks.com/hadoop-tutorial/how-to-install-and-configure-the-hortonworks-odbc-driver-on-mac-os-x/