



NYU

**TANDON SCHOOL
OF ENGINEERING**

Microsoft Azure Cloud & Hortonworks Hadoop

4/30/19

1



NEW YORK UNIVERSITY

Leading invention, innovation
and entrepreneurship



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/>**

← → ⌂ ⌂ https://azure.microsoft.com/en-us/free/ ⌂ ⌂ ⌂ ⌂ ⌂ ⌂ ⌂ ⌂ ⌂ ⌂ ⌂ ⌂ ⌂ ⌂

 Microsoft Azure Contact Sales: 1-800-867-1389 Search My account Portal Sign in

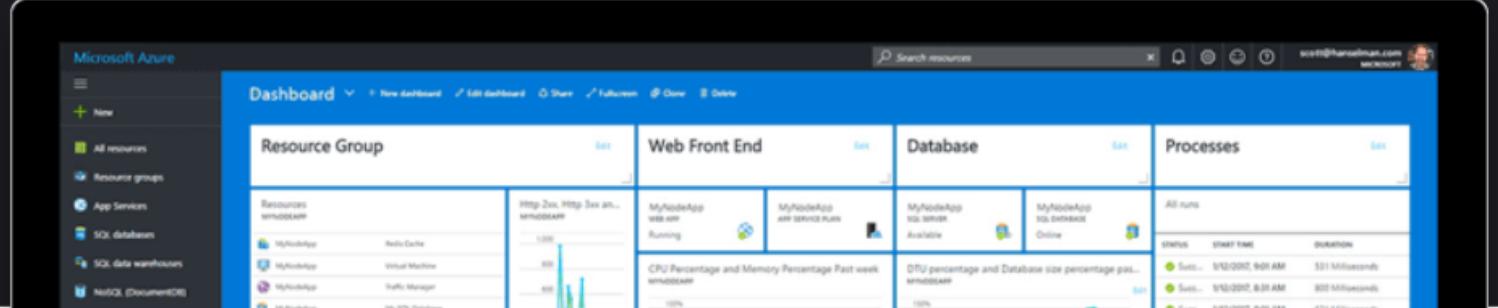
Overview Solutions Products Documentation Pricing Training Marketplace Partners Support Blog More

Create your Azure tree account today

Get started with 12 months of free services

[Start free >](#)

[Or buy now >](#)



3

NYU·poly

POLYTECHNIC INSTITUTE OF NEW YORK UNIVERSITY



Azure services

See all (100+) >

Create a resource >

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>

Virtual machines



App Services



Storage accounts



SQL databases



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

<

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 Server

[Linux Virtual Machines](#)

Provision Ubuntu, Red Hat, CentOS, Ma

[App Service](#)

Create Web Apps using .NET, Java, M

[Functions](#)

Process events with a serverless cod

[SQL Database](#)

Managed relational SQL Database as



Service Health



Marketplace

NYU Poly

POLYTECHNIC INSTITUTE OF NEW YORK UNIVERSITY

Navigate to the ‘MarketPlace’

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

Dashboard > Get Started > Hortonworks Data Platform (HDP) Sandbox

Hortonworks Data Platform (HDP) Sandbox

Hortonworks



Hortonworks Data Platform (HDP) Sandbox

Hortonworks

Create

Save for later

Want to deploy programmatically? Get started →

About To Deploy?

For a step-by-step guide on how to deploy the Hortonworks Sandbox on Azure, visit: [Deploying Hortonworks Sandbox on Microsoft Azure](#).

Already Set Up and Looking to Learn?

There are a series of tutorials to get you going with HDP fast. To learn more about the HDP Sandbox check out: [Learning the Ropes of the Hortonworks HDP Sandbox](#). To get started using Hadoop to store, process and query data try this HDP 2.6 tutorial series: [Hello HDP an introduction to Hadoop](#)

Click Create to Create a Virtual Machine

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

Create a virtual machine that runs Linux or Windows. Select an image from Azure marketplace or use your own customized image.

Complete the Basics tab then Review + create to provision a virtual machine with default parameters or review each tab for full customization.

Looking for classic VMs? [Create VM from Azure Marketplace](#)

PROJECT DETAILS

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

* Subscription i

Free Trial v

 * Resource group i

Select existing... v

[Create new](#)

INSTANCE DETAILS

* Virtual machine name i

[Review + create](#)

[Previous](#)

[Next : Disks >](#)

7



POLYTECHNIC INSTITUTE OF NEW YORK UNIVERSITY

Create VM (Virtual Machine) for Hortonworks 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

« Dashboard > Marketplace > Everything > Hortonworks Sandbox with HDP 2.6.4 > Create a virtual machine

Create a virtual machine

[Basics](#) [Disks](#) [Networking](#) [Management](#) [Guest config](#) [Tags](#) [Review + create](#)

Create a virtual machine that runs Linux or Windows. Select an image from Azure marketplace or use your own customized image. Complete the Basics tab then Review + create to provision a virtual machine with default parameters or review each tab for full customization.

Looking for classic VMs? [Create VM from Azure Marketplace](#)

PROJECT DETAILS

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

* Subscription [Free Trial](#)

* Resource group [\(New\) sandboxst290](#)
[Create new](#)

INSTANCE DETAILS

* Virtual machine name [sandboxst290](#)

* Region [East US](#)

Availability options [No infrastructure redundancy required](#)

* Image [Hortonworks Sandbox with HDP 2.6.4](#)
[Browse all images and disks](#)

[Review + create](#) [Previous](#) [Next : Disks >](#)

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

The screenshot shows the 'Create a virtual machine' wizard in the Azure portal. The left sidebar lists various services: Create a resource, Dashboard, All services, Favorites, All resources, Resource groups, App Services, Function Apps, SQL databases, Azure Cosmos DB, Virtual machines, Load balancers, Storage accounts, Virtual networks, Azure Active Directory, Monitor, Advisor, Security Center, and Cost Management + Bill... The main page title is 'Create a virtual machine'. The 'Size' field is set to 'Standard A6' (4 vcpus, 28 GB memory), with a 'Change size' link. The 'Authentication type' is set to 'Password', selected by a radio button. The 'Username' field contains 'st290'. The 'Password' and 'Confirm password' fields both contain '*****'. Under 'INBOUND PORT RULES', the 'Public inbound ports' dropdown is set to 'None'. A note states: 'All traffic from the internet will be blocked by default. You will be able to change inbound port rules in the VM > Networking page.' Navigation buttons at the bottom include 'Review + create', 'Previous', and 'Next : Disks >'.

Dashboard > Marketplace > Everything > Hortonworks Sandbox with HDP 2.6.4 > Create a virtual machine

Create a virtual machine

* Size ⓘ

Standard A6
4 vcpus, 28 GB memory
[Change size](#)

ADMINISTRATOR ACCOUNT

Authentication type ⓘ

Password SSH public key

* Username ⓘ

st290

* Password ⓘ

* Confirm password ⓘ

INBOUND PORT RULES

Select which virtual machine network ports are accessible from the public internet. You can specify more limited or granular network access on the Networking tab.

* Public inbound ports ⓘ

None Allow selected ports

Select inbound ports

Select one or more ports

i All traffic from the internet will be blocked by default. You will be able to change inbound port rules in the VM > Networking page.

[Review + create](#) [Previous](#) [Next : Disks >](#)

Select Virtual Machine Disk Type: Standard SSD

Dashboard > Marketplace > Everything > Hortonworks Sandbox with HDP 2.6.4 > Create a virtual machine

Create a virtual machine

Basics Disks Networking Management Guest config Tags Review + create

Azure VMs have one operating system disk and a temporary disk for short-term storage. You can attach additional data disks. The size of the VM determines the type of storage you can use and the number of data disks allowed. [Learn more](#)

DISK OPTIONS

* OS disk type [?](#) Standard SSD [▼](#)

DATA DISKS

You can add and configure additional data disks for your virtual machine or attach existing disks. This VM also comes with a temporary disk.

LUN	NAME	SIZE (GiB)	DISK TYPE	HOST CACHING
-----	------	------------	-----------	--------------

[Create and attach a new disk](#) [Attach an existing disk](#)

[▼ ADVANCED](#)

[Review + create](#) [Previous](#) [Next : Networking >](#)

Verify your VM network setting similar to the following

Dashboard > Marketplace > Everything > Hortonworks Sandbox with HDP 2.6.4 > Create a virtual machine

Create a virtual machine

Basics Disks Networking Management Guest config Tags Review + create

Configure a new or existing virtual network for your VM as well as how your VM will be accessed on the virtual network. [Learn more](#)

NETWORK INTERFACE

When creating a virtual machine, a network interface will be created for you.

* Virtual network [i](#) (new) sandboxst290-vnet [Create new](#)

* Subnet [i](#) (new) default (10.0.1.0/24)

Public IP [i](#) (new) sandboxst290-ip [Create new](#)

Network security group Basic Advanced

* Public inbound ports [i](#) None Allow selected ports

Select inbound ports Select one or more ports

Accelerated networking [i](#) On Off

i All traffic from the internet will be blocked by default. You will be able to change inbound port rules in the VM > Networking page.

[Review + create](#) [Previous](#) [Next : Management >](#)

Verify your VM management setting similar to the following

The screenshot shows the Azure portal interface for creating a virtual machine. The left sidebar contains navigation links for various services like Create a resource, Dashboard, All services, Favorites, All resources, Resource groups, App Services, Function Apps, SQL databases, Azure Cosmos DB, Virtual machines, Load balancers, Storage accounts, Virtual networks, Azure Active Directory, Monitor, Advisor, Security Center, and Cost Management + Bill... The main content area is titled 'Create a virtual machine' and shows the 'Management' tab selected. The 'Management' tab allows configuring monitoring and management options for the VM. Under 'MONITORING', 'Boot diagnostics' is set to 'On' and 'OS guest diagnostics' is set to 'Off'. A dropdown for 'Diagnostics storage account' shows '(new) sandboxst290diag' selected, with a 'Create new' link below it. Under 'IDENTITY', 'System assigned managed identity' is set to 'Off'. Under 'AUTO-SHUTDOWN', 'Enable auto-shutdown' is set to 'On', and the 'Shutdown time' is set to '4:00:00 AM' in the '(UTC-05:00) Eastern Time (US & Canada)' time zone. 'Notification before shutdown' is set to 'Off'. At the bottom, there are buttons for 'Review + create', 'Previous', and 'Next : Guest config >'.

Dashboard > Marketplace > Everything > Hortonworks Sandbox with HDP 2.6.4 > Create a virtual machine

Create a virtual machine

Basics Disks Networking Management Guest config Tags Review + create

Configure monitoring and management options for your VM.

MONITORING

Boot diagnostics On Off

OS guest diagnostics On Off

* Diagnostics storage account

IDENTITY

System assigned managed identity On Off

AUTO-SHUTDOWN

Enable auto-shutdown On Off

Shutdown time

Time zone

Notification before shutdown On Off

Review + create **Previous** **Next : Guest config >**

-  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

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!

The screenshot shows the Azure portal interface for creating a virtual machine. The left sidebar contains a navigation menu with various services like Create a resource, Dashboard, All services, Favorites, All resources, Resource groups, App Services, Function Apps, SQL databases, Azure Cosmos DB, Virtual machines, Load balancers, Storage accounts, Virtual networks, Azure Active Directory, Monitor, Advisor, Security Center, and Cost Management + Bill... A green box highlights the 'Favorites' section.

The main content area shows the 'Create a virtual machine' wizard. The breadcrumb navigation at the top is: Dashboard > Marketplace > Everything > Hortonworks Sandbox with HDP 2.6.4 > Create a virtual machine. Below the breadcrumb, the title 'Create a virtual machine' is displayed. A horizontal navigation bar includes tabs for Basics, Disks, Networking, Management, Guest config, Tags (which is underlined), and Review + create.

The 'Tags' section contains the following text: 'Tags are name/value pairs that enable you to categorize resources and view consolidated billing by applying the same tag to multiple resources and resource groups.' It also includes a note: 'Note that if you create tags and then change resource settings on other tabs, your tags will be automatically updated.'

A table is present for defining tags:

KEY	VALUE	RESOURCE TYPE
		All resources to be created

At the bottom of the page are buttons for 'Review + create' (in blue), 'Previous', and 'Next : Review + create >'.

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

Dashboard > Marketplace > Get Started > Hortonworks Data Platform (HDP) Sandbox > Create a virtual machine

Create a virtual machine

 Validation passed

Basics Disks Networking Management Advanced Tags Review + create

PRODUCT DETAILS

Hortonworks Data Platform (HDP) Sandbox by Hortonworks [Terms of use](#) | [Privacy policy](#)

Not covered by credits ⓘ **0.0000 USD/hr**

Standard A6 by Microsoft [Terms of use](#) | [Privacy policy](#)

Subscription credits apply ⓘ **0.5000 USD/hr**

Pricing for other VM sizes

TERMS

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. See the

[Create](#)

[Previous](#)

[Next](#)

[Download a template for automation](#)

15



NEW YORK UNIVERSITY

4/30/19



POLYTECHNIC INSTITUTE OF NEW YORK UNIVERSITY

Leading invention, innovation
and entrepreneurship



Your VM is in the process of deployment

Dashboard > CreateVm-hortonworks.hortonworks-sandbox-sandbox2-20181201201059 - Overview

CreateVm-hortonworks.hortonworks-sandbox-sandbox2-20181201201059 - Overview

Deployment

Search (Ctrl+ /) « Delete Cancel Redeploy Refresh

Overview Inputs Outputs Template

... Your deployment is underway

Check the status of your deployment, manage resources, or troubleshoot deployment issues. Pin this page to your dashboard to easily find it next time.

 Deployment name: CreateVm-hortonworks.hortonworks-sandbox-sandbox2-20181201201059
Subscription: Pay-As-You-Go
Resource group: st290Fall2018

DEPLOYMENT DETAILS [\(Download\)](#)

Start time: 12/1/2018, 8:22:52 PM
Duration: 18 seconds
Correlation ID: 730b67e9-b9d1-457b-b533-43e5c6f7e6a5

RESOURCE	TYPE	STATUS	OPERATION DETAILS
 st290Fall2018-vnet	Microsoft.Network/...	Created	Operation details
 st290IP	Microsoft.Network/...	OK	Operation details
 st290fall2018diag	Microsoft.Storage/s...	Accepted	Operation details
 st290VM-nsg	Microsoft.Network/...	Created	Operation details

16

Wait until your VM deployment is complete

Dashboard > CreateVm-hortonworks.hortonworks-sandbox-sandbox2-20181201201059 - Overview

CreateVm-hortonworks.hortonworks-sandbox-sandbox2-20181201201059 - Overview

Deployment

 Delete  Cancel  Redeploy  Refresh

 Overview

 Inputs

 Outputs

 Template

✓ Your deployment is complete

[Go to resource](#)



Deployment name: CreateVm-hortonworks.hortonworks-sandbox-sandbox2-20181201201059

Subscription: Pay-As-You-Go

Resource group: st290Fall2018

DEPLOYMENT DETAILS [\(Download\)](#)

Start time: 12/1/2018, 8:22:52 PM

Duration: 2 minutes 40 seconds

Correlation ID: 730b67e9-b9d1-457b-b533-43e5c6f7e6a5

RESOURCE	TYPE	STATUS	OPERATION DETAILS
 shutdown-computevr	Microsoft.DevTestL...	Created	Operation details
 st290VM	Microsoft.Compute...	OK	Operation details
 st290vm267	Microsoft.Network/...	Created	Operation details
 st290Fall2018-vnet	Microsoft.Network/...	OK	Operation details
 st290IP	Microsoft.Network/...	OK	Operation details
 st290fall2018diag	Microsoft.Storage/s...	OK	Operation details
 vnetdiag	Microsoft.Network/...	OK	Operation details

Additional Resources



Windows Server
2016 VM
Quickstart
tutorial



Cosmos DB
Quickstart
tutorial



Web App
Quickstart
tutorial



SQL Database
Quickstart
tutorial



Storage Account
Quickstart
tutorial

Helpful Links

[Get started with Azure](#) 
[Azure architecture center](#) 

Your Microsoft Azure Virtual Machine is ready!

Home > sandboxst290

 **sandboxst290**
Virtual machine

Connect Start Restart Stop Capture Move Delete Refresh

Resource group [\(change\)](#)
sandboxst290

Status
Running

Location
East US

Subscription [\(change\)](#)
Pay-As-You-Go

Subscription ID
088e5d98-18c3-461a-95f5-96c15a6cb557

Computer name
sandboxst290

Operating system
Linux

Size
Standard A4 (8 vcpus, 14 GB memory)

Public IP address
13.82.28.75

Virtual network/subnet
sandboxst290-vnet/default

DNS name
[Configure](#)

SETTINGS

 Networking

 Disks

 Size

 Security (Preview)

 Extensions

 Availability set

 Configuration

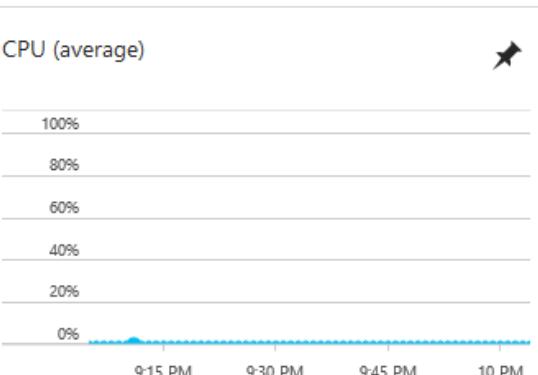
 Properties

 Locks

 Automation script

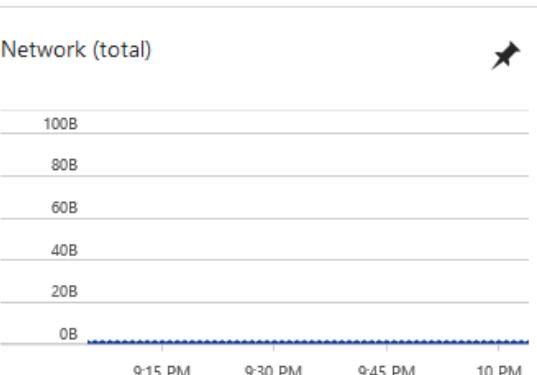
Show data for last: 1 hour 6 hours 12 hours 1 day 7 days 30 days

CPU (average)



PERCENTAGE CPU 0 %

Network (total)



NETWORK IN 0 B NETWORK OUT 0 B

Change your Public IP address to Static (Click Save)

Home > sandboxst290 > sandboxst290-ip - Configuration

sandboxst290-ip - Configuration

Public IP address

Save Discard

The associated virtual machine 'sandboxst290' may be rebooted. [Click here to learn more.](#)

Assignment

Dynamic Static

IP address [?](#)
13.82.28.75

Idle timeout (minutes) [?](#)
 4

DNS name label (optional) [?](#)

Prefer to use your own domain name? [Try Azure DNS now](#)

Configuration Properties Locks Automation script

New support request

4/30/19

NEW YORK UNIVERSITY POLYTECHNIC INSTITUTE OF NEW YORK UNIVERSITY

Leading invention, innovation and entrepreneurship

Open firewall for access your VM

- Click All resources and select Network Security Group

NAME	TYPE	RESOURCE GROUP	LOCATION	SUBSCRIPTION	...
sandboxst290	Disk	SANDBOXST290	East US	Pay-As-You-Go	...
sandboxst290	Virtual machine	sandboxst290	East US	Pay-As-You-Go	...
sandboxst290582	Network interface	sandboxst290	East US	Pay-As-You-Go	...
sandboxst290diag951	Storage account	sandboxst290	East US	Pay-As-You-Go	...
sandboxst290-ip	Public IP address	sandboxst290	East US	Pay-As-You-Go	...
sandboxst290-nsg	Network security gro...	sandboxst290	East US	Pay-As-You-Go	...
sandboxst290-vnet	Virtual network	sandboxst290	East US	Pay-As-You-Go	...

20



Select Network Security Group

Home > All resources > sandboxst290-nsg

sandboxst290-nsg

Network security group

Search (Ctrl+)

Move Delete

Resource group (change)
sandboxst290

Location
East US

Subscription (change)
Pay-As-You-Go

Subscription ID
088e5d98-18c3-461a-95f5-96c15a6cb557

Associated with
0 subnets, 1 network interfaces

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Inbound security rules

PRIORITY	NAME	PORT	PROTOCOL	SOURCE	DESTINATION	ACTION
1000	⚠ default-allow-ssh	22	TCP	Any	Any	Allow
65000	AllowVnetInBound	Any	Any	VirtualNetwork	VirtualNetwork	Allow
65001	AllowAzureLoadBalancerInBo...	Any	Any	AzureLoadBala...	Any	Allow
65500	DenyAllInBound	Any	Any	Any	Any	Deny

Outbound security rules

PRIORITY	NAME	PORT	PROTOCOL	SOURCE	DESTINATION	ACTION
65000	AllowVnetOutBound	Any	Any	VirtualNetwork	VirtualNetwork	Allow
65001	AllowInternetOutBound	Any	Any	Any	Internet	Allow
65500	DenyAllOutBound	Any	Any	Any	Any	Deny

SETTINGS

Inbound security rules

Outbound security rules

Network interfaces

Subnets

Properties

Locks

Automation script

MONITORING

Diagnostics logs

SUPPORT + TROUBLESHOOTING

NI

4/50/10

21

ation and entrepreneurship

I2E

Add an inbound rule to open firewall of your VM

- Click Inbound Security rules and then click Add

The screenshot shows the Azure portal interface for managing Network Security Groups (NSGs). The left sidebar shows navigation options like Overview, Activity log, Access control (IAM), Tags, and Diagnose and solve problems. Under SETTINGS, 'Inbound security rules' is selected. The main content area displays a table of existing rules:

PRIORITY	NAME	PORT	PROTOCOL	SOURCE	DESTINATION	ACTION
1000	default-allow-ssh	22	TCP	Any	Any	Allow
65000	AllowVnetInBound	Any	Any	VirtualNetwork	VirtualNetwork	Allow
65001	AllowAzureLoadBalancerInBo...	Any	Any	AzureLoadBala...	Any	Allow
65500	DenyAllInBound	Any	Any	Any	Any	Deny

At the top right, there are 'Add' and 'Default rules' buttons.

22



POLYTECHNIC INSTITUTE OF NEW YORK UNIVERSITY

sandboxst290-nsg - Inbound security rules

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

Dashboard > All resources > sandboxst290-nsg - Inbound security rules

sandboxst290-nsg - Inbound security rules

Network security group

« + Add Default rules

PRIORITY	NAME	PORT	PROTOCOL	SOURCE	DESTINA...	ACTION
100	HueAccess	1024-6...	Any	Any	Any	Allow ...
65000	AllowVnetInBou...	Any	Any	Virtual...	Virtual...	Allow ...
65001	AllowAzureLoad...	Any	Any	AzureL...	Any	Allow ...
65500	DenyAllInBound	Any	Any	Any	Any	Deny ...

24



POLYTECHNIC INSTITUTE OF NEW YORK UNIVERSITY

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

The screenshot shows a web browser window with the URL 13.82.28.75:8888. The page title is "Hortonworks Sandbox with HDP". The main content area features the Hortonworks logo and the text "SAND BOX HDP2.6.4". Below this, there are two sections: "NEW TO HDP" and "ADVANCED HDP". The "NEW TO HDP" section includes an "HDP" logo icon and a "LAUNCH DASHBOARD" button. The "ADVANCED HDP" section includes a gear icon and a "QUICK LINKS" button. A "GET HELP" button is located in the top right corner of the main content area.

Inbound security rules - Microsoft Edge Hortonworks Sandbox with HDP +

13.82.28.75:8888

Most Visited Getting Started NYU Email

Search

GET HELP

SAND BOX HDP2.6.4

NEW TO HDP

Explore the Hortonworks Data Platform (HDP)

Walk through a typical use case with the tutorial

LAUNCH DASHBOARD

ADVANCED HDP

Expand your Hortonworks Data Platform (HDP) experience

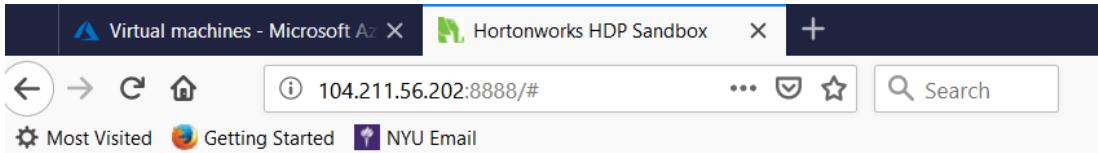
Access components in Sandbox

QUICK LINKS

NEW

25

Launch Dashboard with Username/Password: raj_ops/raj_ops



SAND  HDP2.6.4



NEW TO HDP

Explore the Hortonworks Data Platform (HDP)

Walk through a typical use case with the tutorial

LAUNCH DASHBOARD

26

Make sure all the processes are running

The screenshot shows the Ambari Dashboard interface. On the left, a sidebar lists various Hadoop services: HDFS, YARN, MapReduce, Tez, Hive, HBase, Pig, Sqoop, Oozie, ZooKeeper, Falcon, Storm, Flume, Ambari Infrastructure, Alias, Kafka, Knox, Ranger, Spark2, Zeppelin, Notebook, and Slider. The main content area displays a table titled "0 Background Operations Running". The table has columns for "Operations", "Start Time", "Duration", and a progress bar. There are 10 entries listed:

Operations	Start Time	Duration	Progress
✓ Start All Services	Today 12:39	902.02 secs	100%
✓ Start HDFS	Today 12:33	329.59 secs	100%
✓ Start HiveServer2	Fri May 04 2018 14:11	63.06 secs	100%
✓ Start HiveServer2	Fri May 04 2018 14:10	71.94 secs	100%
✓ Restart all components for Oozie	Fri May 04 2018 14:08	99.83 secs	100%
✓ Start All Services	Fri May 04 2018 13:53	912.12 secs	100%
✓ Start HDFS	Fri May 04 2018 13:48	336.54 secs	100%
— Start All Services	Fri May 04 2018 13:39	417.53 secs	100%
✓ Start HDFS	Fri May 04 2018 13:34	294.56 secs	100%
✓ Start All Services	Tue May 01 2018 20:53	845.54 secs	100%

At the bottom of the dashboard, there are four circular performance metrics: 20%, 602.8 s, 0%, and 1/1. A large green "OK" button is visible on the right.

27



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

104.211.56.202:8080/#/main/dashboard/metrics

67% ...

Ambari Sandbox 0 ops 0 alerts Dashboard Services Hosts Alerts Admin raj_ops

HDFS YARN MapReduce2 Tez Hive HBase Pig Sqoop Oozie ZooKeeper Falcon Storm Flume Ambari Infra Atlas Kafka Knox Ranger Spark2 Zeppelin Notebook Slider Actions

Metrics Heatmaps Config History Metric Actions Last 1 hour

HDFS Disk Usage  47%	DataNodes Live 1/1	HDFS Links NameNode Secondary NameNode 1 DataNodes More...	Memory Usage No Data Available	Network Usage No Data Available
CPU Usage No Data Available	Cluster Load No Data Available	NameNode Heap  20%	NameNode RPC 0.50 ms	NameNode CPU WIO n/a
NameNode Uptime 2.00 hours Fri Dec 07 2018 09:55:46 	HBase Master Heap n/a	HBase Links No Active Master 1 RegionServers n/a More...	HBase Ave Load n/a	HBase Master Uptime n/a
ResourceManager Heap  30%	ResourceManager Uptime 1.8 hr	YARN Memory  0%	NodeManagers Live 1/1	YARN Links ResourceManager 1 NodeManagers More...
Supervisors Live 0/1	Flume Live 1/1			

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.

Inbound security rules - Microsoft Problem loading page

13.82.28.75:8000

Most Visited Getting Started NYU Email

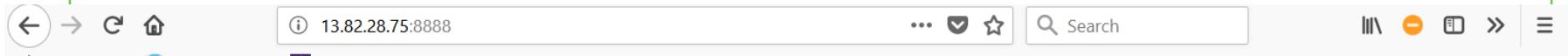
Unable to connect

Firefox can't establish a connection to the server at 13.82.28.75:8000.

- The site could be temporarily unavailable or too busy. Try again in a few moments.
- If you are unable to load any pages, check your computer's network connection.
- If your computer or network is protected by a firewall or proxy, make sure that Firefox is permitted to access the Web.

Try Again

Click Quick Links for Advanced HDP

[GET HELP](#)

SAND BOX HDP2.6.4



NEW TO HDP

[Explore the Hortonworks Data Platform \(HDP\)](#)

Walk through a typical use case with the tutorial

[LAUNCH DASHBOARD](#)

ADVANCED HDP

[Expand your Hortonworks Data Platform \(HDP\) experience](#)

Access components in Sandbox

[QUICK LINKS](#)

50

POLYTECHNIC INSTITUTE OF NEW YORK UNIVERSITY

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

The screenshot shows a web browser window with three tabs:

- Inbound security rules - Microsoft Edge
- Hortonworks Sandbox with HD
- root@sandbox-hdp:~ - Shell In

The address bar displays the URL `13.82.28.75:4200`. Below the address bar are browser navigation icons (back, forward, search, etc.) and a menu icon.

The main content area is a terminal window showing the following session:

```
sandbox login: root
root@sandbox.hortonworks.com's password:
You are required to change your password immediately (root enforced)
Last login: Thu Mar 22 02:26:43 2018 from 127.0.0.1
Changing password for root.
(current) UNIX password:
New password:
Retype new password:
[root@sandbox-hdp ~]#
```

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.
```

```
#####
# 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
# 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=secretkeysecretkeysecretkeysecretkeysecretkey
```

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
```



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]#
```



NEV

Start Hue

Inbound security ru X Hortonworks Sand X root@sandbox-hdp X root@sandbox-hdp X New Tab X

← → ⌂ ⌂ 13.82.28.75:4200 ... ⌂ ⌂ ⌂ Search

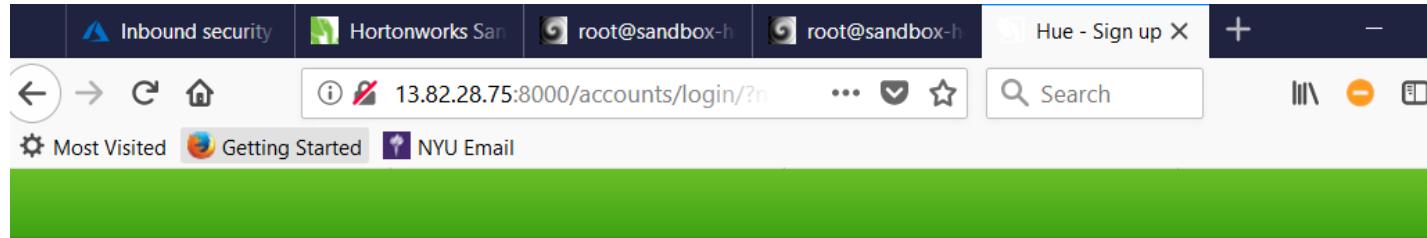
Most Visited Getting Started NYU Email

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



Username

This field is required.

Password

This field is required.

⚠ Since this is your first time logging in, pick any username and password. Be sure to remember these, as they will become your Hue superuser credentials..

13.82.28.75:8000/about/

Most Visited Getting Started NYU Email

Configuration Check for misconfiguration Server details Server Logs

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.



File Brower

Search for file name

A Rename M Move C Copy

+ New + Upload

Change Permissions Download Delete

Home / user / hue Trash

Type	Name	Size	User	Group	Permissions	Date
Folder	.		hue	hue	drwxr-xr-x	March 21, 2018 07:51 PM
Folder	..		hdfs	hdfs	drwxr-xr-x	March 21, 2018 07:51 PM

A screenshot of a web browser window. The address bar shows the URL `13.82.28.75:8000/filebrowser/`. The title bar includes tabs for "Inbound security", "Hortonworks San", "root@sandbox-h", "root@sandbox-h", "File Browser", and a "+" button. Below the address bar are standard navigation buttons (back, forward, home) and a search bar with the placeholder "Search". Underneath the address bar is a toolbar with icons for "Most Visited", "Getting Started", and "NYU Email". A dark navigation bar at the bottom contains icons for Home, User, Hue, and Edit.

Uploading to: /user/hue

The file will then be extracted in the path specified above.

Upload a zip file

File Brow

Search for file name

Change Permiss

Home / user / hue

Inbound security Hortonworks San root@sandbox-h root@sandbox-h File Browser X +

← → ⌂ ⌂ 13.82.28.75:8000/filebrowser/ ... ⌂ ⌂ Search

Most Visited Getting Started NYU Email

File Brow

Search for file name

Change Permiss

Home / user / hue

Upload a zip file

stocks.zip 6% from 0.1GB Cancel

Type	Name	Size	User	Group	Permissions
Folder	.		hue	hue	drwxr-xr-x
Folder			hdbs	hdbs	drwxr-xr-x

4/30/19 POLYTECHNIC INSTITUTE OF NEW YORK UNIVERSITY

Leading invention, innovation and entrepreneurship

File Loading is complete

The screenshot shows a web browser window for the Hortonworks Data Platform File Browser. The URL is 13.82.28.75:8000/filebrowser/view/user/hue. The browser interface includes a top navigation bar with tabs for Inbound security, Hortonworks San, root@sandbox-h, and root@sandbox-h. Below the tabs is a toolbar with icons for back, forward, search, and other file operations. A green header bar contains links for Most Visited, Getting Started, NYU Email, and several application icons. The main content area is titled "File Browser" and shows the "hue" directory structure. It includes a toolbar with "Search for file name", "Rename", "Move", "Copy", "New", "Upload", "Change Permissions", "Download", and "Delete". The directory listing shows three items: a hidden directory "..", the "stocks" directory, and another hidden directory "..". All files and directories belong to the "hue" user and group, with permissions set to drwxr-xr-x. The last modified date for all items is March 21, 2018, at 07:59 PM.

Type	Name	Size	User	Group	Permissions	Date
..	.		hue	hue	drwxr-xr-x	March 21, 2018 07:59 PM
..	..		hdfs	hdfs	drwxr-xr-x	March 21, 2018 07:51 PM
stocks	stocks		hue	hue	drwxr-xr-x	March 21, 2018 07:59 PM

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 screenshot shows the Hue web interface. At the top, there's a browser-style header with back, forward, search, and other navigation buttons. Below it is a toolbar with various icons (e.g., HDFS, HCat, File Browser) and a user dropdown labeled 'hue'. The main navigation bar includes 'Query Editor' (which is highlighted in blue), 'My Queries', 'Saved Queries', 'History', 'Result', 'Databases', 'Tables', 'Settings', and 'Invalidate session'. The central area is titled 'Query Editor' and contains a code editor window. On the left of the code editor is a sidebar with sections for 'DATABASE' (set to 'default'), 'SETTINGS' (with an 'Add' button), 'FILE RESOURCES' (with an 'Add' button), 'USER-DEFINED FUNCTIONS' (with an 'Add' button), 'PARAMETERIZATION' (with a checked 'Enable' checkbox and a 'Parameterization' link), and 'EMAIL NOTIFICATION' (with an unchecked 'Email me on completion' checkbox). The code editor itself has two lines of DDL:

```
1 create external table price_data (stock_exchange string, symbol string,  
2
```

At the bottom of the code editor are buttons for 'Execute', 'Save as...', 'Explain', and 'New query'.

Query Editor

```
1 create external table price_data (stock_exchange string, symbol string,  
2
```

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'

The screenshot shows the Hue Query Editor interface. At the top, there's a green header bar with various icons and a user dropdown. Below it is a navigation bar with tabs: 'Query Editor' (which is selected), 'My Queries', 'Saved Queries', 'History', 'Result', 'Databases', 'Tables', and 'Settings'. On the far right of the navigation bar is an 'Invalidate session' button. The main area is titled 'Query Results: Unsaved Query'. It has four tabs: 'Results' (selected), 'Query' (which is highlighted with a dashed border), 'Log', and 'Columns'. The 'Query' tab contains the SQL command: 'select * from price_data where symbol = 'IBM''. To the left of the main area, there's a sidebar titled 'DOWNLOADS' with options: 'Download as CSV', 'Download as XLSX', and 'Save'. At the bottom left, there's a yellow callout box with the text: '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.' The background of the entire interface is light gray.



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'

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: 0(+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  
INFO : Map 1: 0(+8)/18 Map 3: 1/1      Reducer 2: 0/8
```

Query Editor My Queries Saved Queries History Result Databases Tables Settings Invalidate session

Query Results: Unsaved Query

DOWNLOADS

Download as CSV

Download as XLSX

Save

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	2
27	IBM	1966	562.0	289.5	2
9	IBM	1967	648.0	362.5	5
28	IBM	1968	649.880004883	304.5	2
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

NEW YORK UNIVERSITY

4/30/19

NYU·poly
POLYTECHNIC INSTITUTE OF NEW YORK UNIVERSITY

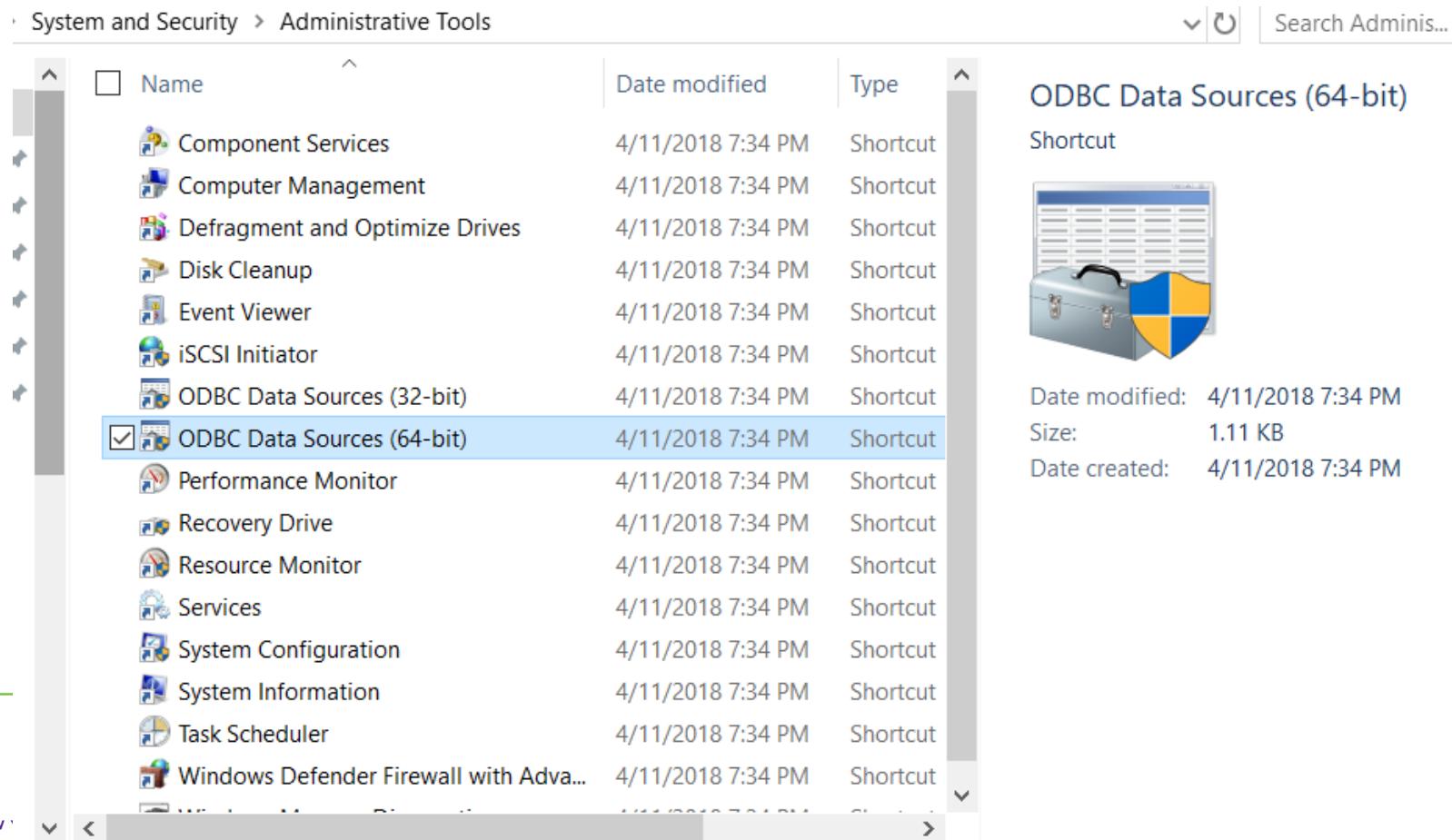
Leading invention, innovation
and entrepreneurship

51

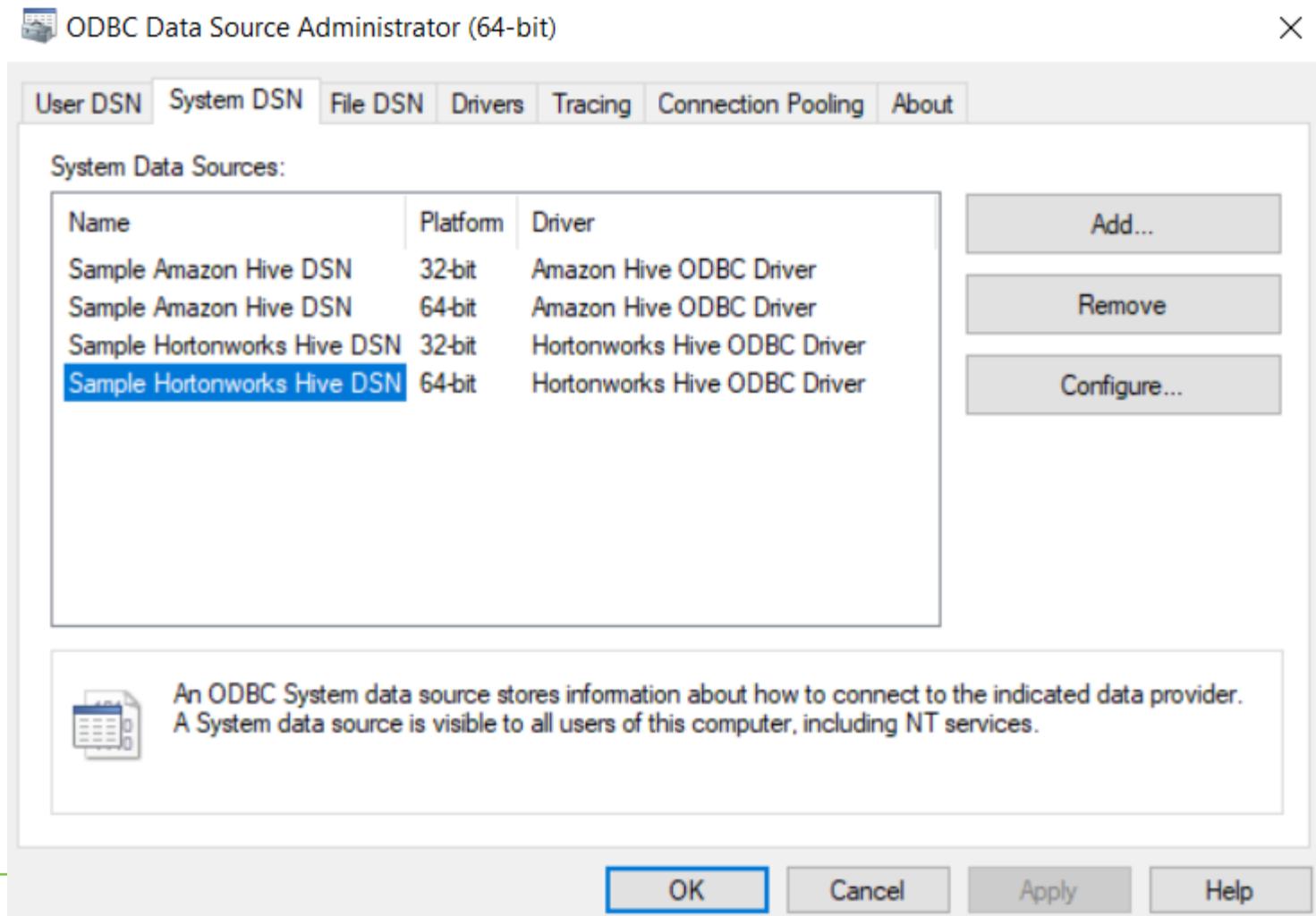
- For Apple Mac Laptops, go to page 57, Instructions of hortonworks Hive ODBC driver installation, and rest procedures.

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.



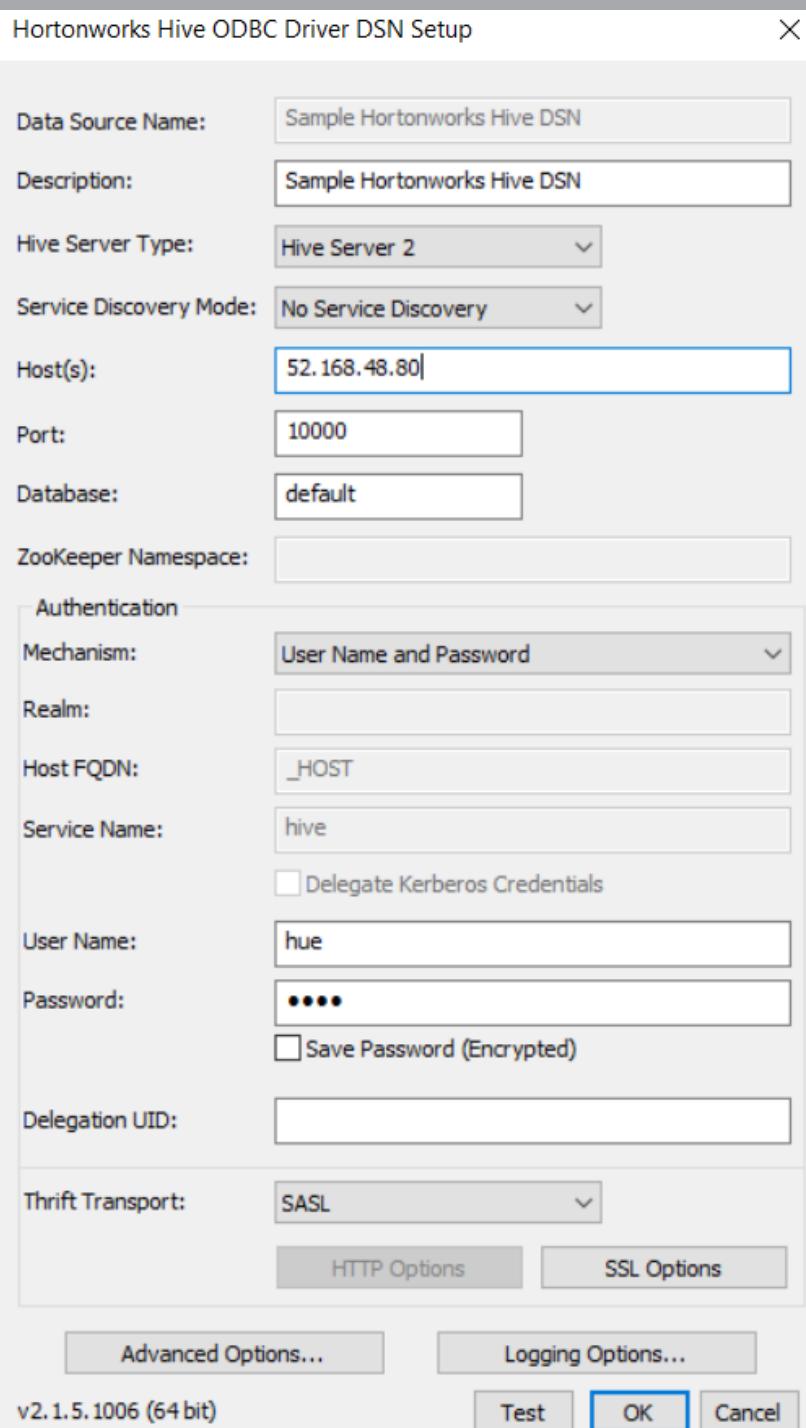
Configure Hortonworks HIVE ODBC Driver



54

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

Test Results

X

Test Results

SUCCESS!

Successfully connected to data source!

ODBC Version: 03.80

Driver Version: 2.1.5.1006

Bitness: 64-bit

Locale: en_US

OK

56

NYU·poly

POLYTECHNIC INSTITUTE OF NEW YORK UNIVERSITY

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();

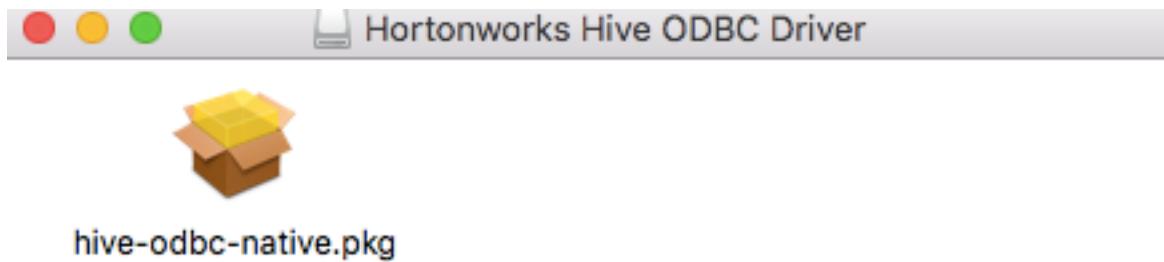
cursor.execute("select symbol,trade_date as dividend_date, dividend from
default.dividends_data where symbol = 'IBM' and year(trade_date) = '1990'")

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

The screenshot shows the 'Security & Privacy' preferences window in macOS. The 'General' tab is selected. At the top, it says 'A login password has been set for this user' with a 'Change Password...' button. Below that are three checkboxes: 'Require password' (checked, dropdown menu set to 'immediately'), 'Show a message when the screen is locked' (unchecked), and 'Disable automatic login' (checked). A horizontal line separates this from the 'Allow apps downloaded from:' section. Underneath, there are two radio buttons: 'App Store' (unchecked) and 'App Store and identified developers' (checked). At the bottom left is a padlock icon with the text 'Click the lock to prevent further changes.' On the bottom right are 'Advanced...' and a help button (?). The window has standard OS X window controls (red, yellow, green buttons) and a search bar at the top right.

Security & Privacy

Search

General FileVault Firewall Privacy

A login password has been set for this user Change Password...

Require password immediately after sleep or screen saver begins

Show a message when the screen is locked Set Lock Message...

Disable automatic login

Allow apps downloaded from:

App Store

App Store and identified developers

Click the lock to prevent further changes.

Advanced... ?



Step 2: Download ODBC configuration file

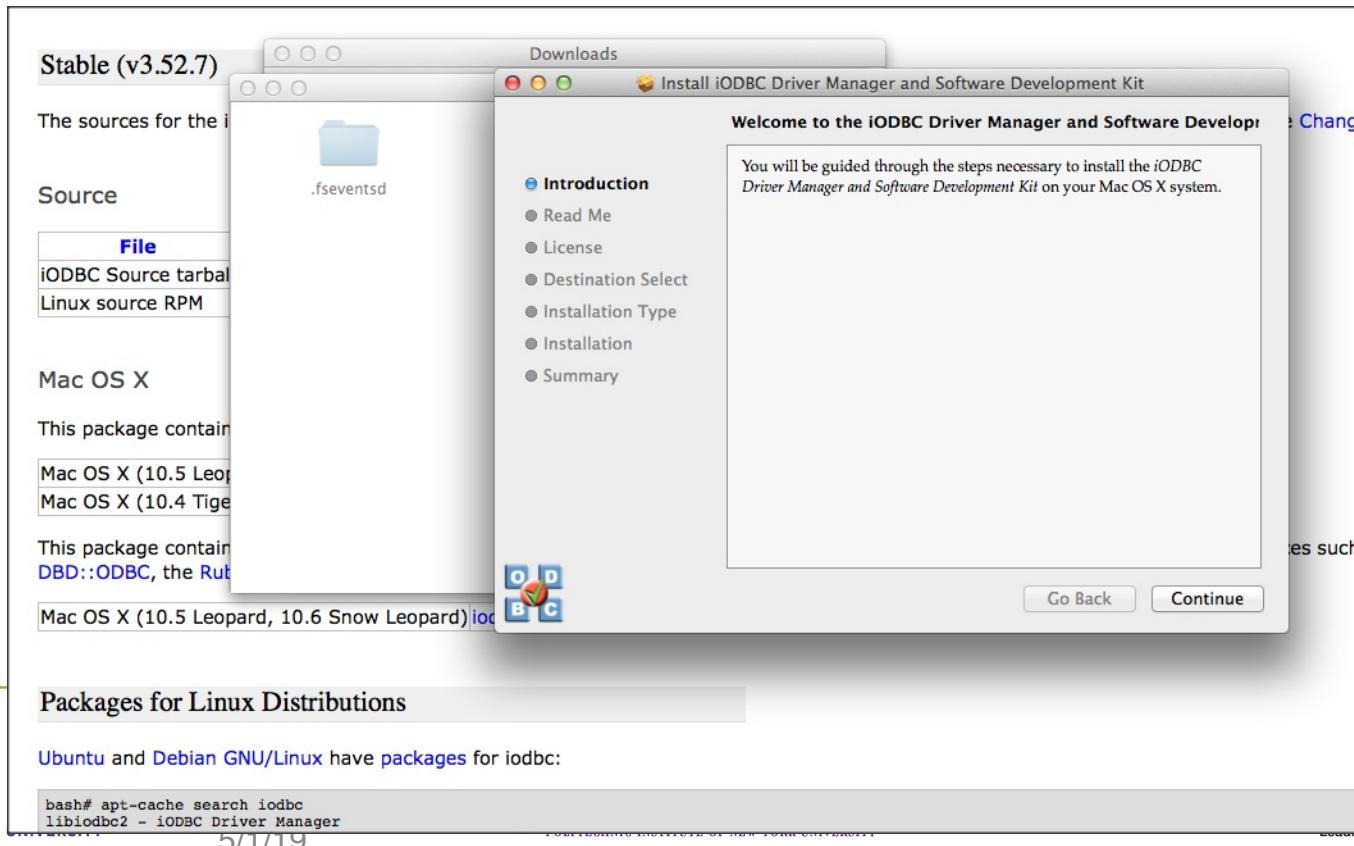
- From ODBC Driver for MAC folder of our class Web site, download the file: odbc.ini
- Run the following commands via a terminal:
 - \$ cd ~
 - \$ cd Library/ODBC
 - \$ sudo mv odbc.ini odbc.ini.bak
 - \$ cd ~
 - \$ cd Downloads/
 - \$ sudo cp odbc.ini ../Library/ODBC/
- Use vi to change the host to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



Stable (v3.52.7)

The sources for the i

Source

File

iODBC Source tarball
Linux source RPM

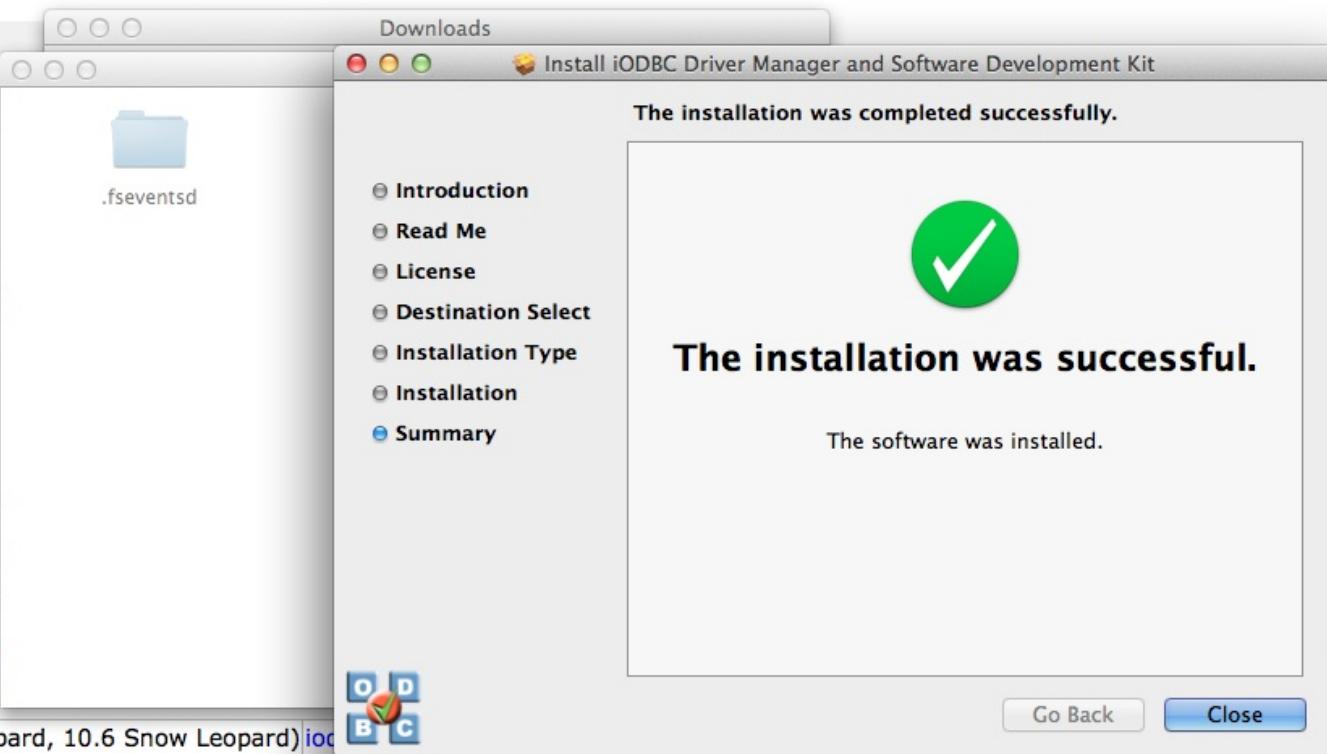
Mac OS X

This package contains

Mac OS X (10.5 Leopard)
Mac OS X (10.4 Tiger)

This package contains
DBD::ODBC, the Ruby

Mac OS X (10.5 Leopard, 10.6 Snow Leopard) iodbc



Packages for Linux Distributions

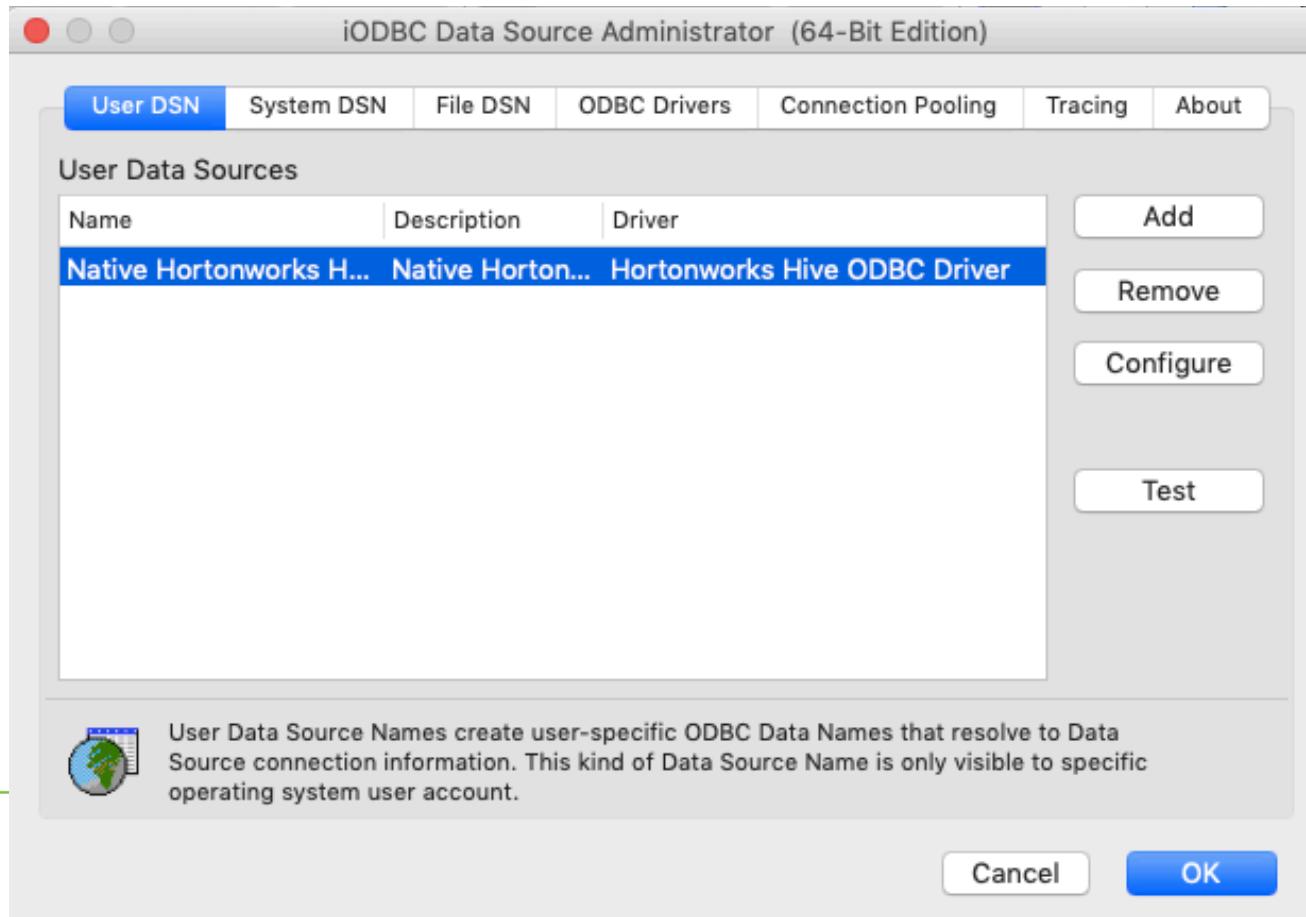
Ubuntu and Debian GNU/Linux have packages for iodbc:

```
bash# apt-cache search iodbc
libiodbc2 - iODBC Driver Manager
libiodbc2-dev - iODBC Driver Manager (development files)
```

64

Test the connection to your Azure Hortonwork

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



65

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

A screenshot of a terminal window titled "root@sandbox-hd" with the command "service hue stop" being run. The terminal output shows the service stopping successfully and then exiting.

```
Inbound security X Hortonworks San X root@sandbox-hd X root@sandbox-hd X + - □ X
← → ⌂ ⌂ 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. 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]# service hue start
Detecting versions of components...
HUE_VERSION=2.6.1-91
Starting hue: [ OK ]
[root@sandbox-hdp hue]# service hue stop
Shutting down hue: [ OK ]
[root@sandbox-hdp hue]# exit
```

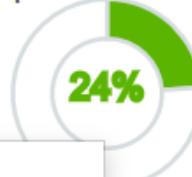
Stop all Hortonworks Processes by Click Stop ALL

104.211.56.202:8080/#/main/dashboard/metrics

- Flume
- Ambari Infra
- Atlas
- Kafka
- Knox
- Ranger
- Spark2
- Zeppelin
- Notebook
- Slider

Actions ▾

- + Add Service
- ▶ Start All
- Stop All
- ⟳ Restart All Required
- ⬇ Download All Client Configs

NameNode Uptime	HBase Master Heap	HBase Links
3.1 hr	n/a	No Active 1 Region n/a
ResourceManager Heap	ResourceManager Uptime	YARN Memo
 24%	2.9 hr	0.0% 0 Bytes o
Flume Live	Flume Live	
1/1	1/1	

All the Hortonworks processes are stopped

0 Background Operations Running

Operations	Start Time	Duration	Show:
✓ Stop All Services	Today 13:05	191.63 secs	<div style="width: 100%; background-color: #2e7131; height: 10px;"></div> 100% ►
✓ Start All Services	Today 09:59	17.11 mins	<div style="width: 100%; background-color: #2e7131; height: 10px;"></div> 100% ►
✓ Start HDFS	Today 09:53	380.85 secs	<div style="width: 100%; background-color: #2e7131; height: 10px;"></div> 100% ►
— Start All Services	Today 09:52	52.95 secs	<div style="width: 100%; background-color: #f0ad4e; height: 10px;"></div> 100% ►
✓ Stop All Services	Thu Dec 06 2018 20:57	329.43 secs	<div style="width: 100%; background-color: #2e7131; height: 10px;"></div> 100% ►

<

OK

70

<

>



Stop your VM by clicking Stop!

Home > sandboxst290

 **sandboxst290**
Virtual machine

Connect Start Restart Stop Capture Move Delete Refresh

Resource group (change)
sandboxst290

Status
Running

Location
East US

Subscription (change)
Pay-As-You-Go

Subscription ID
088e5d98-18c3-461a-95f5-96c15a6cb557

Computer name
sandboxst290

Operating system
Linux

Size
Standard A4 (8 vcpus, 14 GB memory)

Public IP address
13.82.28.75

Virtual network/subnet
sandboxst290-vnet/default

DNS name
Configure

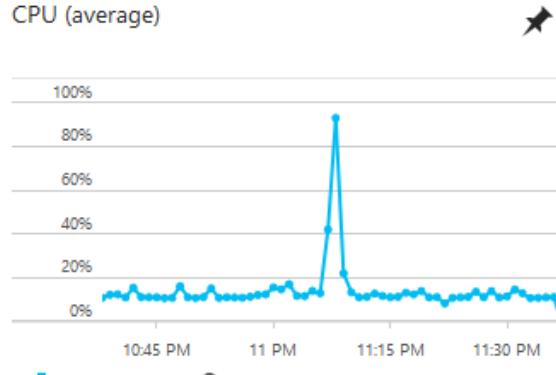
Overview Activity log Access control (IAM) Tags Diagnose and solve problems

SETTINGS

Networking Disks Size Security (Preview) Extensions Availability set Configuration Properties

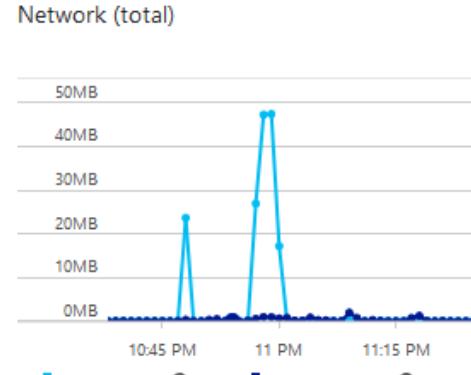
Show data for last: **1 hour** 6 hours 12 hours 1 day 7 days 30 days

CPU (average)



100%
80%
60%
40%
20%
0%
10:45 PM 11 PM 11:15 PM 11:30 PM
PERCENTAGE CPU

Network (total)



50MB
40MB
30MB
20MB
10MB
0MB
10:45 PM 11 PM 11:15 PM 11:30 PM
NETWORK IN
NETWORK OUT

71 1



Verify your VM is stopped

Home > Virtual machines

Virtual machines

Default Directory

Add Reservations Edit columns Refresh Assign tags Start Restart Stop Delete

Subscriptions: Pay-As-You-Go

Filter by name... All resource groups All types All locations All tag

2 items

<input type="checkbox"/> NAME ↑↓	TYPE ↑↓	STATUS	RESOURCE GROUP ↑↓	LOCATION ↑↓
<input type="checkbox"/>  sandboxst290	Virtual machine	Stopped (dealloc...)	sandboxst290	East US
<input type="checkbox"/>  st290VM	Virtual machine	Stopped (dealloc...)	st290Fall2018	East US

72



POLYTECHNIC INSTITUTE OF NEW YORK UNIVERSITY

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/>