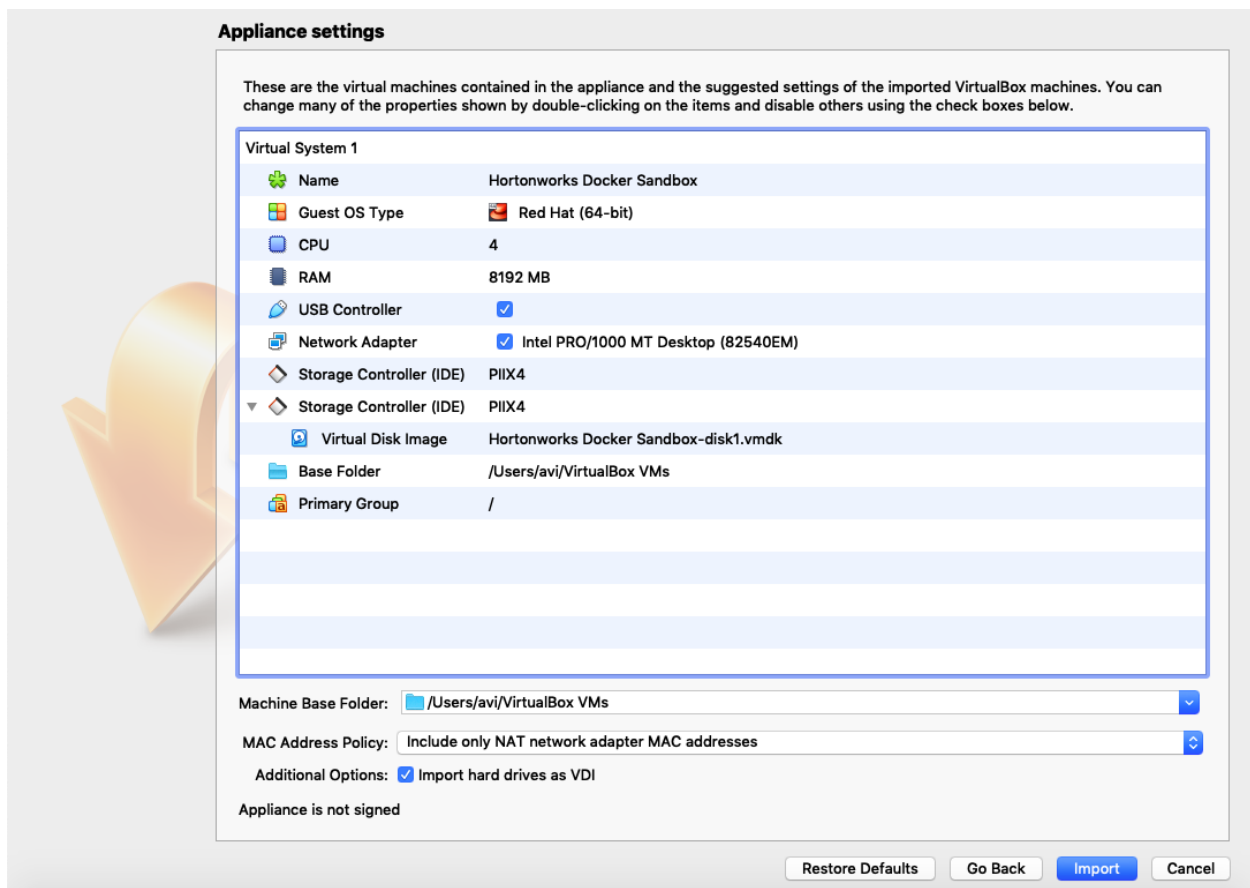
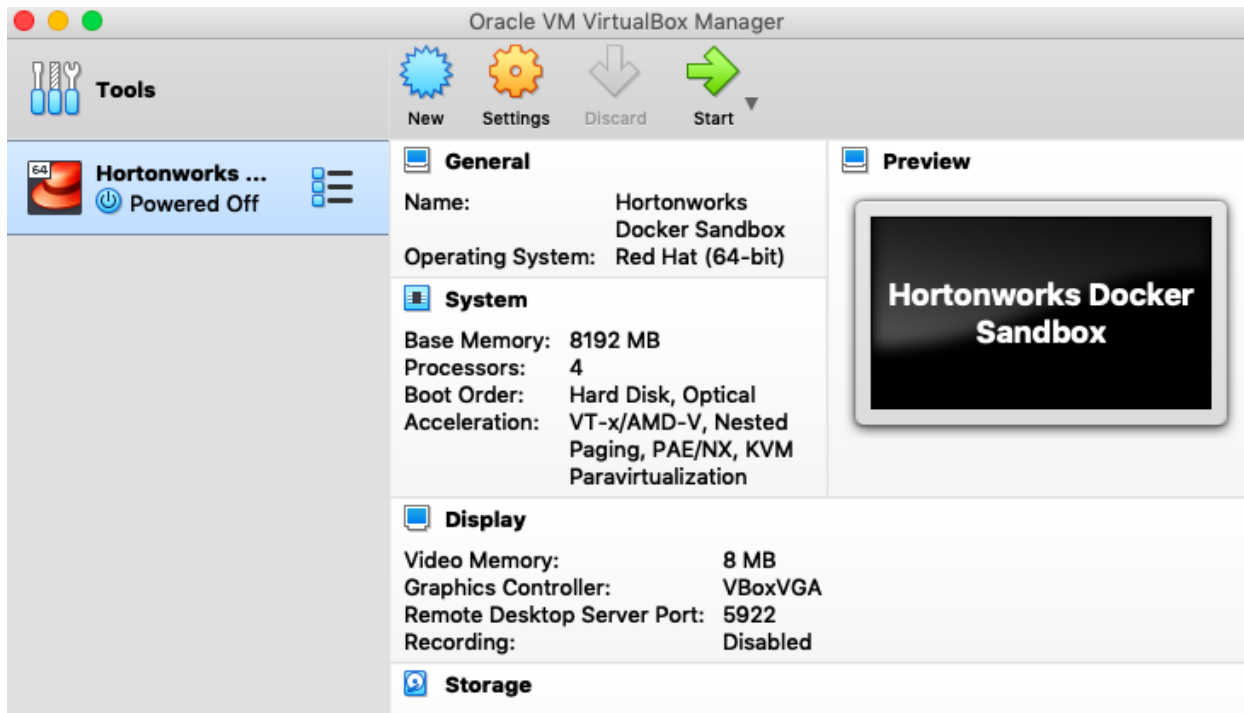


CLOUD COMPUTING ASSIGNMENT 3

**Avinash Ganguri
16293133**

Step1: Downloaded the Virtual box hosted hypervisor for Mac



Step2: Download Hortonworks HDP Image from the site,

The screenshot shows the Cloudera website's download page for Hortonworks Data Platform (HDP) on the Hortonworks Sandbox. The page features a navigation bar with links to 'Why Cloudera', 'Products', 'Solutions', and 'Services & Support'. The main content area has a large heading 'Hortonworks Data Platform (HDP*) on Hortonworks Sandbox' and a subheading 'Get Started Now'. Below this, there is a dropdown menu for 'Installation Type' set to 'Virtualbox' and a prominent 'LET'S GO!' button. A secondary section titled 'HDF on Sandbox' provides links for 'Download HDF', 'Release Notes', and 'Getting Started with HDP'. A small chatbot icon is visible in the bottom right corner.

CLUSTERA Why Cloudera Products Solutions Services & Support

Hortonworks Data Platform (HDP*) on Hortonworks Sandbox

The HDP Sandbox makes it easy to get started with Apache Hadoop, Apache Spark, Apache Hive, Apache HBase, Druid and Data Analytics Studio (DAS).

Get Started Now

Installation Type
Virtualbox

LET'S GO! →

HDP on Sandbox

Getting Started with HDP >
Learning the Ropes of the HDP Sandbox >
Release Notes >

HDF on Sandbox

Download HDF >
Release Notes >

Hi! Do you want to say "yes" to any analytic workload? 🤖

The screenshot shows a terminal window titled 'Hortonworks Docker Sandbox [Running]'. The terminal displays the following text:

```
HDP 2.5
http://hortonworks.com
```

To initiate your Hortonworks Sandbox session, please open a browser and enter this address in the browser's address field:
`http://127.0.0.1:8888/`

Log in to this virtual machine: Linux/Windows <Alt+F5>, Mac OS X <Fn+Alt+F5>

The terminal window includes a standard macOS-style title bar and a dock at the bottom with various application icons.

And my system configurations are,



*To run the Image Hortonworks HDP in Virtual Box, we need **RAM** more than 8GB. On running the virtual box my system crashes for every 5 mins, so I have to choose another alternative for running Hortonworks HDP Image. Therefore, I used **Microsoft Azure services** to create a Virtual Machine and run the **Hortonworks HDP 2.5 Version** in it.*

Assignment 3 (Due 4/14 12:30) x Home - Microsoft Azure x Ambari - Sandbox x +

portal.azure.com/#home

Apps YouTube Netflix Prime Video India Amazon Prime US GitHub Google Tech Dev... Python Sheet Canvas Cloud Class Python Class

Microsoft Azure Search resources, services, and docs (G+)

agv8x@mail.umkc.edu UNIVERSITY OF MISSOURI

Azure services

Create a resource

Virtual machines Subscriptions Cost Management... Storage accounts SQL databases App Services Azure Database for PostgreSQL Azure Cosmos DB More services

Recent resources

Name Type Last Viewed

Name	Type	Last Viewed
Assignment3	Virtual machine	48 minutes ago
Azure subscription	Subscription	an hour ago
cloud-ip	Public IP address	a month ago
CC-ip	Public IP address	a month ago
cseastus10032	Storage account	a month ago
cloud-shell-stc	Resource group	a month ago

Virtual machines

+ Create View

Recent resources

Assignment3 48 minutes ago

Free training from Microsoft

Introduction to Azure virtual machines 8 units - 1 hr 7 min

Create a Windows virtual machine in Azure 9 units - 51 min

Create a Linux virtual machine in Azure 7 units - 1 hr 26 min

Useful links

Overview Get Started Pricing

Navigate

Subscriptions

Tools

Microsoft Learn Azure Monitor Security Center Cost Management

Secure your apps and Analyze and optimize your

https://portal.azure.com/#blade/HubsExtension/BrowseResourceBlade/resourceType/Microsoft.Compute%2FVirtualMachines

Assignment 3 (Due 4/14 12:30) x Select an image - Microsoft Azure x Ambari - Sandbox x Cloudera Enterprise Download x +

portal.azure.com/#create/Microsoft.VirtualMachine

Apps YouTube Netflix Prime Video India Amazon Prime US GitHub Google Tech Dev... Python Sheet Canvas Cloud Class Python Class

Microsoft Azure Search resources, services, and docs (G+)

agv8x@mail.umkc.edu UNIVERSITY OF MISSOURI

Create a virtual machine

resource group

Create

Instance details

Virtual machine name *

Region * (US)

Availability options * No inf

Image * Ubuntu

Azure Spot instance * No

Size * Standard 2 vcp

Administrator account

Authentication type * Password

Username *

SSH public key *

Review + create < Previous

Select an image

Marketplace My Items

AI + Machine Learning

Analytics

Blockchain

Compute

Containers

Databases

Developer Tools

DevOps

Identity

Integration

Internet of Things

IT & Management Tools

Media

Mixed Reality

Networking

Security

Software as a Service (SaaS)

Storage

Web

Search: horton

Hortonworks Sandbox with HDP 2.6.4

Hortonworks

Powered by HDP 2.6 100% open source platform for Hadoop, Spark, Storm, HBase, Kafka, Hive, Ambari

SlashDB Enterprise Edition

vt.enterprise

Request any number of databases and user accounts; high-touch setup and support

SlashDB Cloud Edition (old)

vt.enterprise

License for production deployments of any size

SlashDB Cloud Edition

vt.enterprise

License for production deployments of any size

Visual Studio Emulator on Windows Server 2016 (V1)

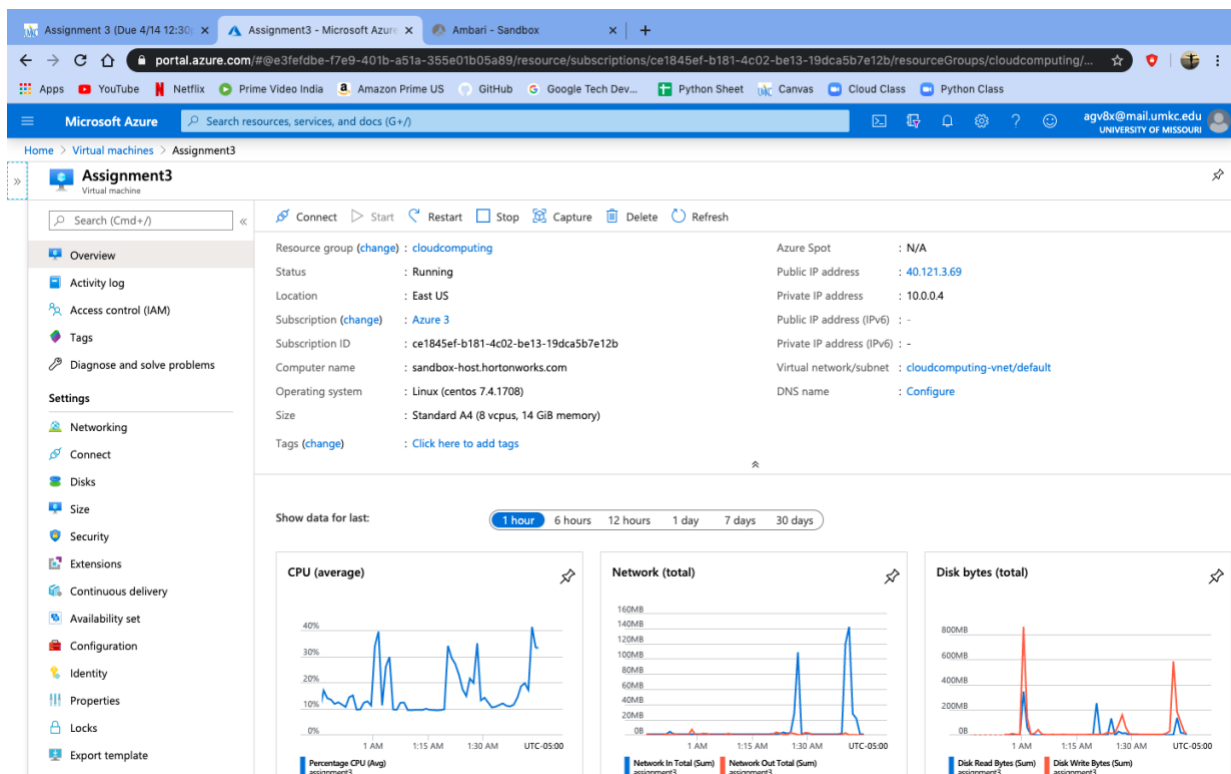
Apps4Rent LLC

Test your Android apps with the best-in-class Android emulator

DataRoad Reflect Version 2.1.0

DataRoad Technologies LLC

Summary: DataRoad Reflect automates data movement quickly and easily with zero coding.

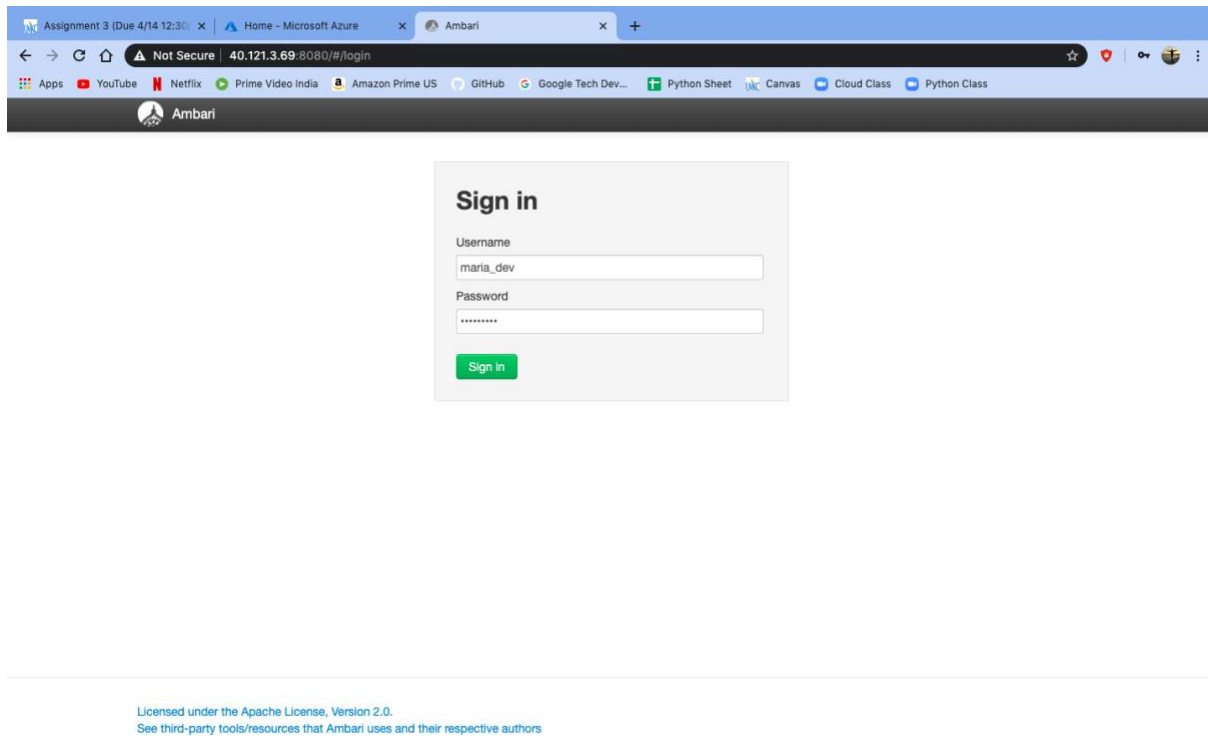


And then connecting the VM to my localhost through SSH,

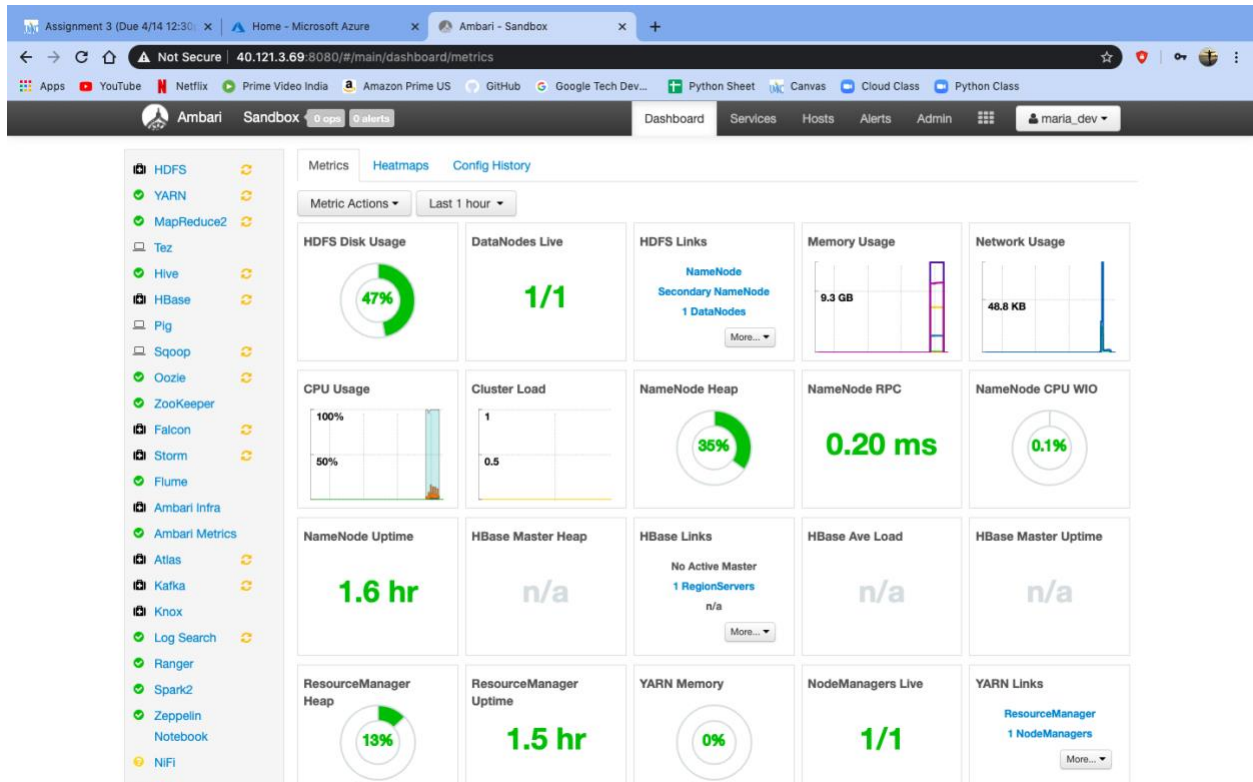
```
avi — avi@sandbox-host:~ — ssh avi@40.121.3.69 — 85x25
Last login: Tue Apr 14 00:30:06 on ttys000
Mac-Air:~ avi$ ssh avi@40.121.3.69
avi@40.121.3.69's password:
Last login: Tue Apr 14 06:01:21 2020 from 136.37.26.194
[avi@sandbox-host ~]$
```

Step 4:

Using the public IP from Azure VM, open it in the browser,
Enter username: maria_dev, password: maria_dev



The Ambari login page shows up, after successful login you can see the dashboard,

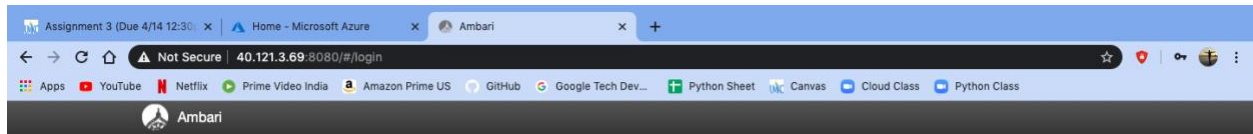


Changing the username and password and also, creating admin credentials by,

```
avi — root@sandbox-hdp:~ — ssh avi@40.121.3.69 — 85x25
Last login: Tue Apr 14 01:00:47 on ttys001
Mac-Air:~ avi$ ssh avi@40.121.3.69
[avi@40.121.3.69's password:
Last login: Tue Apr 14 06:46:50 2020 from 136.34.124.143
[avi@sandbox-host ~]$ ssh -p2222 root@127.0.0.1
root@127.0.0.1's password:
Last login: Tue Apr 14 06:47:45 2020 from 172.17.0.1
[root@sandbox-hdp ~]# ambari-admin-password-reset
Please set the password for admin:
Please retype the password for admin:

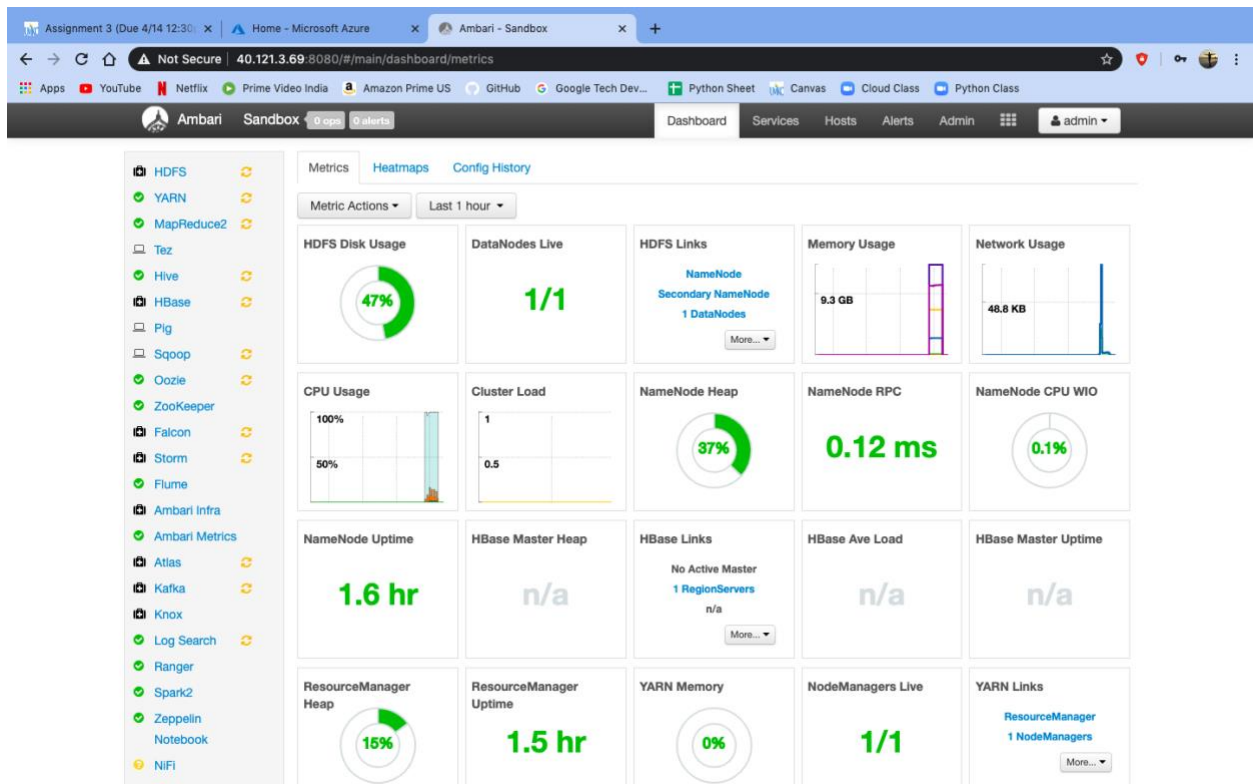
The admin password has been set.
Restarting ambari-server to make the password change effective...

Using python /usr/bin/python
Restarting ambari-server
Waiting for server stop...
Ambari Server stopped
Ambari Server running with administrator privileges.
Organizing resource files at /var/lib/ambari-server/resources...
Ambari database consistency check started...
Server PID at: /var/run/ambari-server/ambari-server.pid
Server out at: /var/log/ambari-server/ambari-server.out
Server log at: /var/log/ambari-server/ambari-server.log
Waiting for server start.....
```

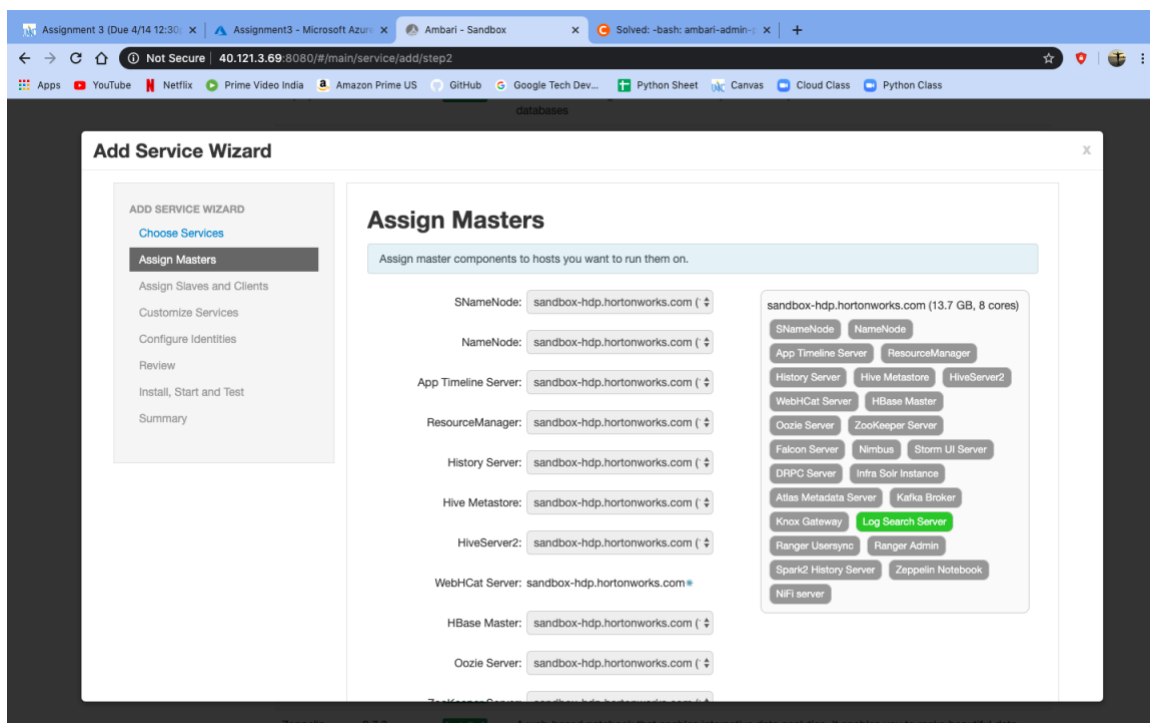
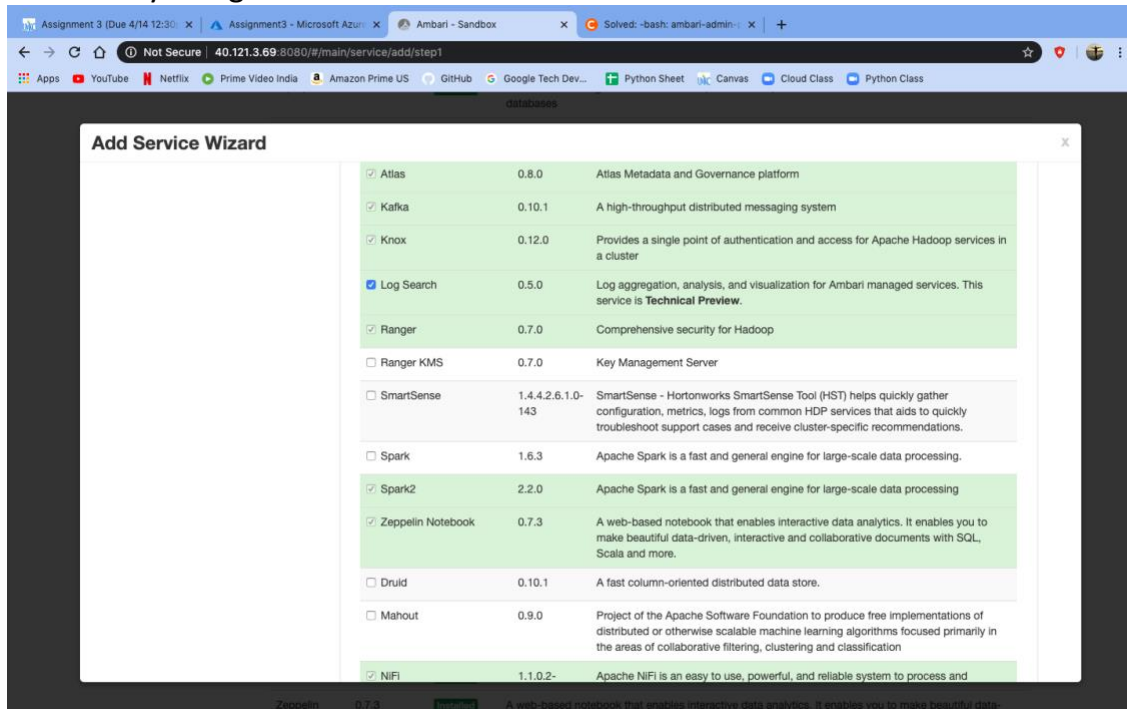
Licensed under the Apache License, Version 2.0.
See third-party tools/resources that Ambari uses and their respective authors

After logging with the admin details, the dashboard page



Step 5: Explore Ambari using web browser. Explore its capabilities and show at least three things with screenshots of its function (eg. installing other modules, setting system alert parameter and method, etc.)

- 1) A Log Search Service is added, and only an Admin can add it, so we have to login only using admin details.



After successfully deploying,

The screenshot shows the 'Add Service Wizard' in the Ambari interface, specifically the 'Customize Services' step. The left sidebar contains a navigation menu with options: Assign Masters, Assign Slaves and Clients, Customize Services (selected), Configure Identities, Review, Install, Start and Test, and Summary. The main content area displays a list of services: HDFS, YARN, MapReduce2, Tez, Hive, HBase, Pig, Sqoop, Oozie, ZooKeeper, Falcon, Storm, Flume, Ambari Infra, Atlas, Kafka, Knox, Log Search (selected), Ranger, Spark2, Zeppelin Notebook, NiFi, Slider, and Misc. Below the service list, there are tabs for 'Settings' and 'Advanced'. The 'Advanced' tab is active, showing configuration for 'logsearch-admin-json'. It includes fields for 'Admin username' (ambari_logsearch_admin) and 'Admin password' (masked with asterisks). A text area for 'logsearch-admin-json template' contains a JSON snippet. At the bottom, there are expandable sections for 'Advanced logfeeder-ambari-config' and 'Advanced logfeeder-custom-logsearch-conf'.

The screenshot shows the 'Add Service Wizard' in the Ambari interface, specifically the 'Install, Start and Test' step. The left sidebar is the same as the previous screenshot, but 'Install, Start and Test' is now selected. The main content area has a heading 'Install, Start and Test' and a message 'Please wait while the selected services are installed and started.' Below this is a progress bar showing '100 % overall'. A table displays the installation status for the selected services:

Host	Status	Message
sandbox-hdp.hortonworks.com	100%	Success

Below the table, it says '1 of 1 hosts showing - Show All' and 'Show: 25 1 - 1 of 1'. A green banner at the bottom states 'Successfully installed and started the services.' with a 'Next -->' button.

At the bottom of the Ambari interface, there is a list of installed services:

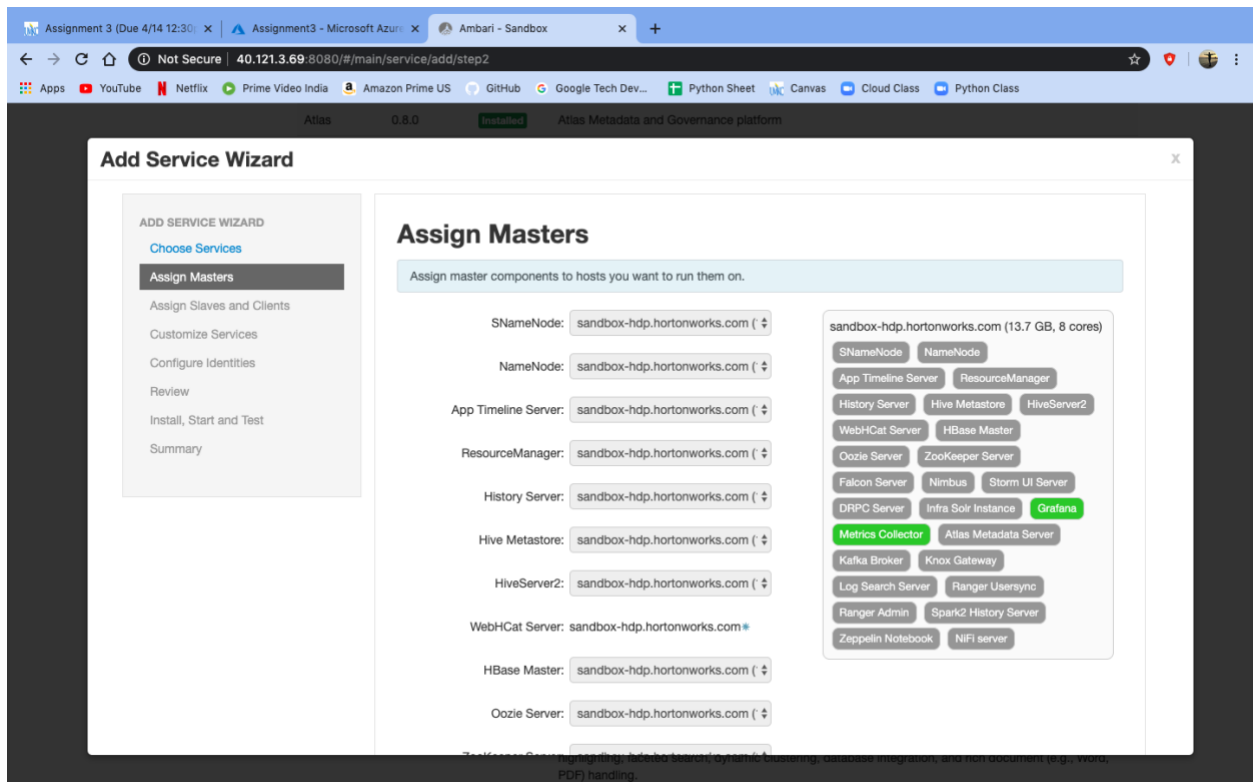
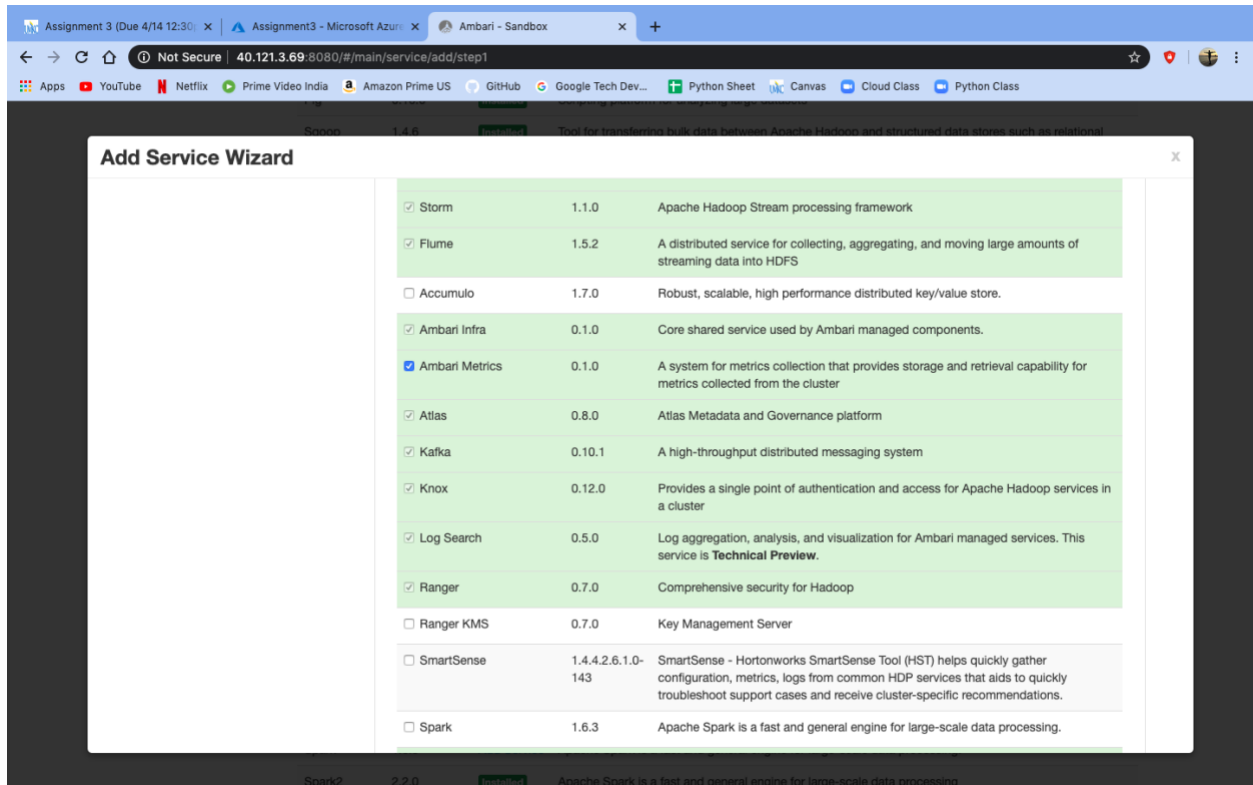
- SmartSense 1.4.4.2.6.1.0-143 Add Service SmartSense - Hortonworks SmartSense Tool (HST) helps quickly gather configuration, metrics, logs from common HDP services that aids to quickly troubleshoot support cases and receive cluster-specific recommendations.
- Spark 1.6.3 Add Service Apache Spark is a fast and general engine for large-scale data processing.
- Spark2 2.2.0 Installed Apache Spark is a fast and general engine for large-scale data processing.
- Zeppelin 0.7.3 Installed A web-based notebook that enables interactive data analytics. It enables you to make beautiful data-

The screenshot shows the Ambari Sandbox interface. On the left, a sidebar lists various services: HDFS, YARN, MapReduce2, Tez, Hive, HBase, Pig, Sqoop, Oozie, ZooKeeper, Falcon, Storm, Flume, Ambari Infra, Atlas, Kafka, Knox, Log Search (highlighted with a red '1' alert icon), Ranger, Spark2, Zeppelin, Notebook, NiFi, and Slider. The main content area shows the 'Log Search' service configuration. The 'Summary' tab is active, displaying a 'Log Search Server' status as 'Started' with a red '1 alert' icon. Below it, a 'Log Feeder' is listed as '1/1 Log Feeder Live'. The top navigation bar includes 'Dashboard', 'Services', 'Hosts', 'Alerts', and 'Admin'. The user is logged in as 'admin'.

2) Adding NameNode Host CPU Utilization Alert by changing the interval from 5 minsto 10 mins and connection timeout 5 secs to 10 secs.

The screenshot shows the Ambari Alerts configuration page for 'NameNode Host CPU Utilization'. The page is titled 'NameNode Host CPU Utilization' with a 'Back' link and a 'NONE' button. The 'Configuration' section includes a description: 'This host-level alert is triggered if CPU utilization of the NameNode exceeds certain warning and critical thresholds. It checks the NameNode JMX Servlet for the SystemCPULoad property. The threshold values are in percent.' The 'Check Interval' is set to '10 Minute'. The 'Thresholds' section shows three levels: 'OK' (green) at '{1} CPU, load (0:1%)', 'WARNING' (orange) at '200 %' with '{1} CPU, load (0:1%)', and 'CRITICAL' (red) at '250 %' with '{1} CPU, load (0:1%)'. The 'Connection Timeout' is set to '10 Seconds' with a 'CRITICAL' label. The 'Instances' section shows a table with columns for 'Service', 'Host', 'Status', '24-Hour', and 'Response'. The table is currently empty, displaying 'No alert instances to display'.

3) Adding Ambari Metrics Service,



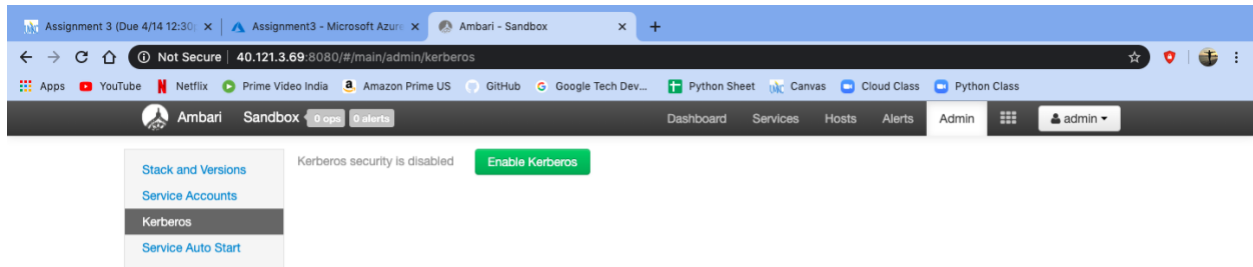
Creating admin password,

The screenshot shows the 'Add Service Wizard' in Ambari, specifically step 4. The wizard is titled 'Add Service Wizard' and has a sidebar with steps: Review, Install, Start and Test, and Summary. The main content area is divided into two sections: 'General' and 'Metric Collector'. In the 'General' section, the 'Metrics Service operation mode' is set to 'embedded'. The 'Metrics Collector log dir' is '/var/log/ambari-metrics-collector', the 'Metrics Collector pid dir' is '/var/run/ambari-metrics-collector', the 'Metrics Monitor log dir' is '/var/log/ambari-metrics-monitor', and the 'Metrics Monitor pid dir' is '/var/run/ambari-metrics-monitor'. The 'Grafana Admin Username' is 'admin', and the 'Grafana Admin Password' is masked with dots. In the 'Metric Collector' section, the 'Metrics service checkpoint delay' is set to 60. The bottom of the screen shows the URL '40.121.3.69:8080/#'.

After successfully deploying,

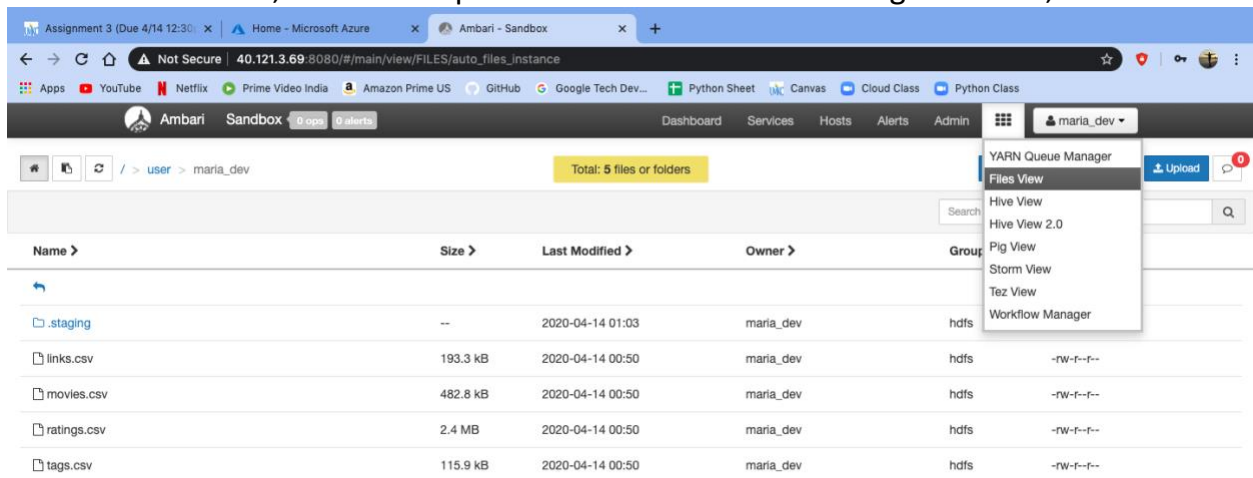
The screenshot shows the 'Add Service Wizard' in Ambari, specifically step 7. The wizard is titled 'Add Service Wizard' and has a sidebar with steps: ADD SERVICE WIZARD, Choose Services, Assign Masters, Assign Slaves and Clients, Customize Services, Configure Identities, Review, Install, Start and Test, and Summary. The main content area is titled 'Install, Start and Test' and shows a progress bar at 100% overall. Below the progress bar, there is a table with columns 'Host', 'Status', and 'Message'. The table shows one host, 'sandbox-hdp.hortonworks.com', with a status of '100%' and a message of 'Success'. The bottom of the screen shows the URL '40.121.3.69:8080/#'.

Host	Status	Message
sandbox-hdp.hortonworks.com	100%	Success



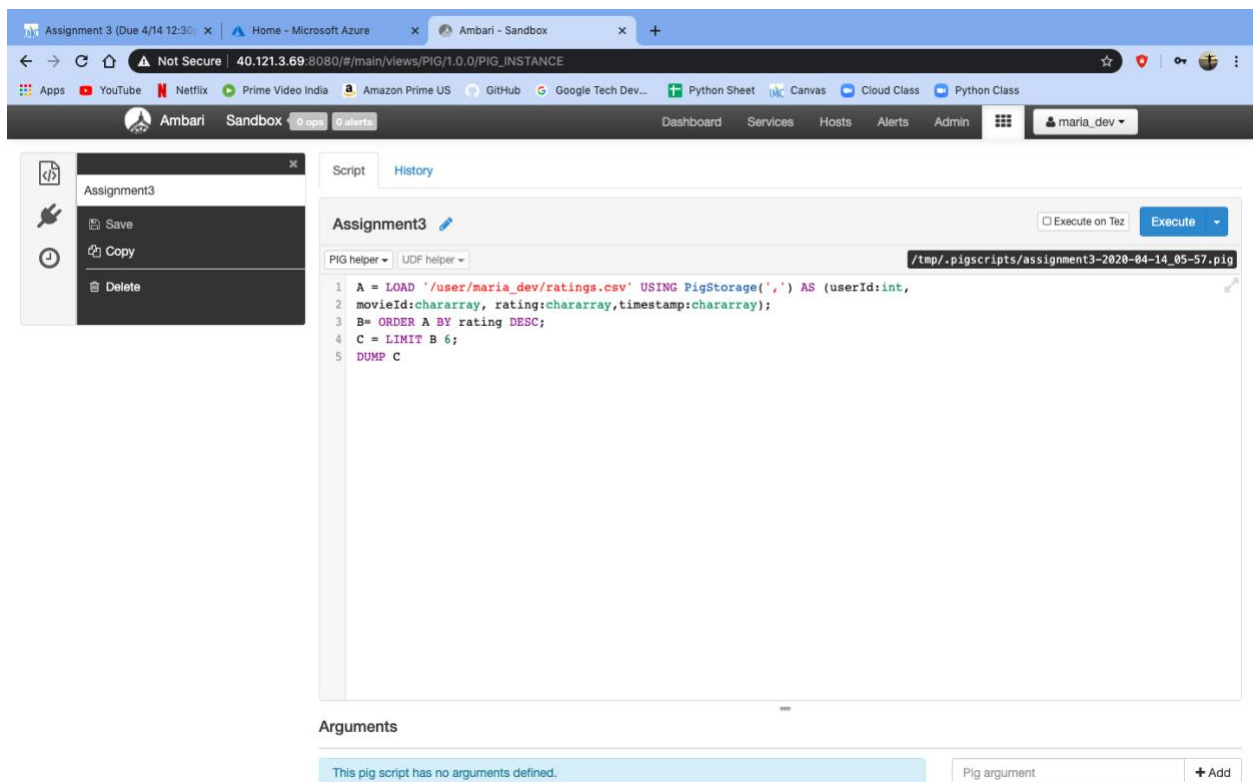
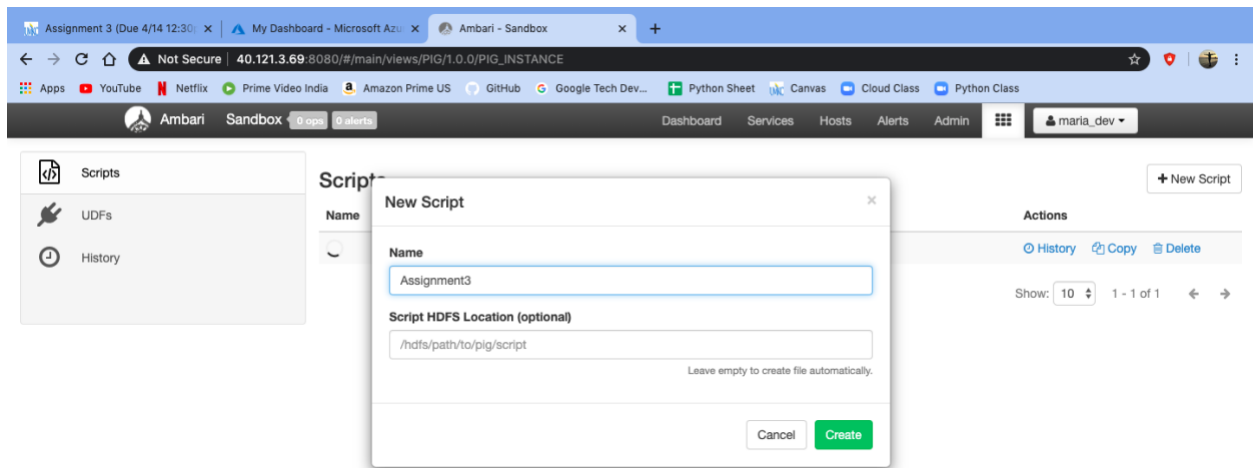
Licensed under the Apache License, Version 2.0.
See third-party tools/resources that Ambari uses and their respective authors

Step 6: To explore Ambari -> Pig View and find top 5 highly rated movies.
This is the file view, where we upload the csv files such as ratings.csv here,



40.121.3.69:8080/#

In the Pig view, a new script is created and we write our code here,



Our script file is executed here and we get the Top 5 results,

The screenshot shows the Ambari Sandbox interface. The top navigation bar includes links to Dashboard, Services, Hosts, Alerts, and Admin. The user is logged in as 'maria_dev'. The main content area displays 'Assignment3 - COMPLETED' with a green progress bar. Below this, the Job ID is 'job_1586841676167_0001' and it started on '2020-04-14 00:58'. The 'Results' section shows the top 5 results of the Pig script execution:

(movieId,rating,timestamp)
(325,1387,5.0,1039398049)
(325,1358,5.0,1039396113)
(325,1248,5.0,1039399628)
(325,1228,5.0,1039395869)
(325,1219,5.0,1039397876)

There are also links for 'Logs' and 'Script Details'.

Step 7: Do the above Step 6 with Tez enabled,

The screenshot shows the Ambari Sandbox interface with the Pig script for 'Assignment3' displayed. The script is as follows:

```
1 A = LOAD '/user/maria_dev/ratings.csv' USING PigStorage(',') AS (userId:int,  
2 movieId:chararray, rating:chararray,timestamp:chararray);  
3 B= ORDER A BY rating DESC;  
4 C = LIMIT B 6;  
5 DUMP C
```

The script is saved in the file '/tmp/.pigscripts/assignment3-2020-04-14_05-57.pig'. Below the script, the 'Arguments' section indicates that no arguments are defined for this Pig script.

Assignment3 - **COMPLETED**

Job ID: job_1586841676167_0003
Started: 2020-04-14 01:02

Results [Download](#)

```
(,movieId,rating,timestamp)
(325,1387,5.0,1039398049)
(325,1358,5.0,1039396113)
(325,1248,5.0,1039399628)
(325,1228,5.0,1039395869)
(325,1219,5.0,1039397876)
```

[Logs](#) [Download](#)

[Script Details](#)

Date	Status	Duration	Actions
2020-04-14 01:02	COMPLETED	1 min, 12 sec	Logs Results Delete
2020-04-14 00:58	COMPLETED	1 min, 55 sec	Logs Results Delete

Show: 10 1 - 2 of 2

With the Tez enabled, the script was executed in 1 min 12 secs and with the Mapreduce it was executed in 1 min 55 sec. Hence, *Tez runs faster than MapReduce*.