Exercise #2:

Introduction to Hortonworks Sandbox

INTRODUCTION

This tutorial is aimed for users who do not have much experience in using the Sandbox. We will install and explore the Sandbox on virtual machine and cloud environments. We will also navigate the Ambari user interface.

Let's begin our Hadoop journey.

PRE-REQUISITES

- Downloaded and Installed Hortonworks Sandbox
- Allow yourself around one hour to complete this tutorial
- If on mac or linux, added sandbox.hortonworks.com to your /private/etc/hosts file
- If on windows 7, added sandbox.hortonworks.com to

YOUr /c/Windows/System32/Drivers/etc/hosts file

If on mac or linux, to add sandbox.hortonworks.com to your list of hosts, open the terminal, enter the following command, replace {Host-Name} with the appropriate host

for your sandbox:

echo '{Host-Name} sandbox.hortonworks.com' | sudo tee -a /etc/hosts

NOTE: In single machine, just replace {Host-Name} with 127.0.0.1

If on windows 7, to add sandbox.hortonworks.com to your list of hosts, open git bash, enter the following command, replace {Host-Name} with the appropriate host for your sandbox:

echo '{Host-Name} sandbox.hortonworks.com' | tee -a
/c/Windows/System32/Drivers/etc/hosts

HW13386:~ mrizvi\$ echo '23.99.9.232 sandbox.hortonworks.com' | sudo tee -a /private/etc/hosts 23.99.9.232 sandbox.hortonworks.com

WHAT IS THE SANDBOX?

The Sandbox is a straightforward, pre-configured, learning environment that contains the latest developments from Apache Hadoop Enterprise, specifically Hortonworks Data Platform (HDP) Distribution. The Sandbox comes packaged in a virtual environment that can run in the cloud or on your personal machine. The Sandbox allows you to learn and explore HDP on your own.

SECTION 1: SANDBOX IN VM

STEP 1: EXPLORE THE SANDBOX IN A VM

1.1 INSTALL THE SANDBOX

Start the Hortonworks Sandbox following the steps in exercise 1 to start the VM.

1.2 LEARN THE HOST ADDRESS OF YOUR ENVIRONMENT

Once you have installed the Sandbox VM, it resolves to the host on your environment. As, a general thumb rule, wait for the installation to complete and confirmation screen will tell you the host your sandbox resolves to. For example:

In case of VirtualBox: host would be 127.0.0.1



1.3 CONNECT TO THE WELCOME SCREEN

Append the port number :8888 to your host address, open your browser, and access

Sandbox Welcome page at http://_host_:8888/.





1.4 MULTIPLE WAYS TO EXECUTE TERMINAL COMMANDS

Note: For all methods below, the login credential instructions will be the same to access the Sandbox through the terminal.

- Login using username as **root** and password as **hadoop**.
- After first time login, you will be prompted to retype your current password, then change your password.
- If you are using Putty on Windows then go to terminal of your sandbox in oracle virtualBox -> Press A1t+F5 -> enter username - root -> enter password

- hadoop -> it will ask you to set new password -> set new password.

Secure Shell (SSH) Method:

Open your terminal (mac and linux) or putty (windows). Type the following command to access the Sandbox through SSH:

```
# Usage:
    ssh <username>@<hostname> -p <port>;
# Example:
    ssh root@127.0.0.1 -p 2222;
```

```
[HW12576:~ jmedel$ ssh root@127.0.0.1 -p 2222;
[root@127.0.0.1's password:
Last login: Mon Feb 29 23:42:16 2016
[root@sandbox ~]#
```

Mac OS Terminal

Shell Web Client Method:

Open your web browser. Type the following text into your browser to access the Sandbox through the shell:

```
# Usage:
    # _host_:4200
Example:
    127.0.0.1:4200
```

```
sandbox login: root
root@sandbox.hortonworks.com's password:
Last login: Mon Feb 29 04:39:02 2016 from 10.0.2.2
[root@sandbox ~]#
```

Appearance of Web Shell

VM Terminal Method:

Open the Sandbox through Virtualbox or VMware. The Sandbox VM Welcome Screen will appear. For Linux/Windows users, press Alt+F5 and for Mac, press Fn+Alt+F5 to login into the Sandbox VM Terminal.

```
Hortonworks Sandbox with HDP 2.4_3 [Running]
```

```
CentOS release 6.7 (Final)
Kernel 2.6.32-573.18.1.el6.x86_64 on an x86_64
To login to the the shell, use:
username: root
password: hadoop
sandbox login: root
Password:
Last login: Mon Feb 29 23:28:11 on tty5
[root@sandbox ~]#_
```

VirtualBox VM Terminal

1.5 LEARN YOUR SANDBOX VERSION

To find information about your sandbox, execute the command:

```
sandbox-version
```

1.6 SEND DATA BETWEEN SANDBOX & LOCAL MACHINE

Open your terminal (linux or mac) or git bash (windows). To send data from your local machine to the sandbox, you would input the following command. If you want to try this command, replace the HDF filename with another filename from your Downloads folder. Modify the command and execute:

scp -P 2222 ~/Downloads/<any-file-of-your-choice> root@localhost:/root

This command sends the file from your local machine's Downloads folder to the Sandbox's root directory. We can send any file, directory we want, we just need to specify the path. We can also choose any sandbox directory or path that we want the data to land into.

Here is the definition of the command that we used above:

```
scp -P <input-port> </input-directory-path-local-machine> <input-username@hostname-
:/sandbox-dir-path>
```

We can also send data from sandbox to our local machine, refer to the modified command definition below:

```
scp -P <input-port> <input-username@hostname-:/sandbox-dir-path> </input-directory-
path-local-mach>
```

What is the difference between the two command definitions above? To send data from local machine to sandbox, the local machine directory path comes before sandbox directory. To transfer data from sandbox to local machine, the command arguments are reversed.

STEP 2: EXPLORE AMBARI

Navigate to Ambari welcome page using the **url** given on Sandbox welcome page. **Note:** Both the username and password to login are **maria_dev**.

2.1 USE TERMINAL TO FIND THE HOST IP SANDBOX RUNS ON

If you want to search for the host address your sandbox is running on, ssh into the sandbox terminal upon successful installation and follow subsequent steps:

- Login using username as root.
- Type ifconfig and look for inet addr: under eth0.
- Use the inet addr, append **:8080** and open it into a browser. It shall direct you to Ambari login page.

• This inet address is randomly generated for every session and therefore differs from session to session.

HW12576	5:~ jmedel\$ ssh root@127.0.0.1 -p 2222						
root@12	27.0.0.1's password:						
Last lo	ogin: Tue Mar 1 03:24:49 2016 from 10.0.2.2						
[root@	sandbox ~]# ifconfig						
eth0	Link encap:Ethernet HWaddr 08:00:27:8C:73:68						
	inet addr:10.0.2.15 Bcast:10.0.2.255 Mask:255.255.255.0						
	UP BROADCAST RUNNING MULTICAST MILL-1500 Metric-1						
	RX packets:2005 errors:0 dropped:0 overrups:0 frame:0						
	TX packets:2009 errors:0 dropped:0 overruns:0 carrier:0						
	collisions: 0 translen: 1000						
	$\frac{1}{1000}$						
	KX Dytes:409042 (400.0 KIB) IX Dytes:2566100 (2.4 MIB)						
	Interrupt:19 Base address:0xd020						
lo	Link encap:Local Loopback						
	inet addr:127.0.0.1 Mask:255.0.0.0						
	UP LOOPBACK RUNNING MTU:65536 Metric:1						
	RX packets:266747 errors:0 dropped:0 overruns:0 frame:0						
	TX nackets: 266747 errors: 0 dropped: 0 overruns: 0 carrier: 0						
	collisions:0 txqueuelen:0						
	PX bytes: 149558996 (142 6 MiR) TX bytes: 149558996 (142 6 MiR)						
	IX bytes.145556556 (142.0 Mtb) IX bytes.145556556 (142.0 Mtb)						

[root@sandbox ~]#

Services Provided By the Sandbox

Service	URL
Sandbox Welcome Pa	age http://host.8888
Ambari Dashboard	http://host.8080
ManageAmbari	http://host:8080/views/ADMIN_VIEW/2.4.0.0/INSTANCE/#/
Hive User View	http://host:8080/#/main/views/HIVE/1.5.0/AUTO_HIVE_INSTANCE
Pig User View	http://host:8080/#/main/views/PIG/1.0.0/Pig_INSTANCE
File User View	http://host:8080/#/main/views/FILES/1.0.0/AUTO_FILES_INSTANCE
SSH Web Client	http:// <i>host</i> :4200
Hadoop Configuration	http://host:50070/dfshealth.html http://host:50070/explorer.html

The following Table Contains Login Credentials:

Service	User	Password
Ambari, OS	admin	refer to step 2.1
Ambari, OS	maria_dev	maria_dev
Ambari, OS	raj_ops	raj_ops
Ambari, OS	holger_gov	holger_gov
Ambari, OS	amy_ds	amy_ds

Please go to Section 3 to know more about these users.

2.2 SETUP AMBARI ADMIN PASSWORD MANUALLY

• Start your sandbox and open a terminal (mac or linux) or putty (windows)

- SSH into the sandbox as root using ssh root@127.0.0.1 -p 2222
- Type the following commands:

```
# Updates password
ambari-admin-password-reset
# If Ambari doesn't restart automatically, restart ambari service
ambari-agent restart
```

Note: Now you can login to ambari as an admin user to perform operations, such as starting and stopping services.



2.3 EXPLORE AMBARI WELCOME SCREEN 5 KEY CAPABILITIES

Enter the Ambari Welcome URL and then you should see a similar screen:

🝌 Ambari	Ambari User Views
Clusters 3 Sandbox © Permissions Go to Dashboard Versions Urews Views	4 Welcome to Apache Ambari Montor your cluster resources, manage who can access the cluster, and customize views for Ambari users.
User + Group Management Users Groups	Manage Users + Groups Wanage the users and groups that can access Amberi Users Groups Create view instances and grant permissions Users Users

- "Operate Your Cluster" will take you to the Ambari Dashboard which is the primary UI for Hadoop Operators
- "Manage Users + Groups" allows you to add & remove Ambari users and groups
- "Clusters" allows you to grant permission to Ambari users and groups
- "Ambari User Views" list the set of Ambari Users views that are part of the cluster
- "Deploy Views" provides administration for adding and removing Ambari User Views

2.4 EXPLORE AMBARI DASHBOARD LINKS

I

Enter the Ambari Dashboard URL and you should see a similar screen:



Click on

- Metrics, Heatmap and Configuration
- and then the
- **Dashboard**, **Services**, **Hosts**, **Alerts**, **Admin** and User Views icon (represented by 3×3 matrix) to become familiar with the Ambari resources available to you.

SECTION 3: NEW USERS IN SANDBOX

Ambari 2.4 introduced the notion of Role-Based Access Control(RBAC) for the Ambari web interface. Ambari now includes additional cluster operation roles providing more granular division of control of the Ambari Dashboard and the various Ambari Views. The image below illustrates the various Ambari Roles. Only the admin id has access to view or change these roles.

Clusters	Sandbox R	oles	
Sandbox 🕑			
Roles			
Go to Dashboard		5 S. 21. 5	BLUCK US
Versions	Roles 😡	Assign roles to these users	Assign roles to these groups
Remote Clusters	Cluster Administrator	raj_ops	Add Group
Views		Q	
Views View URLs	Cluster Operator		Add Group
LUser + Group Management	Service Administrator	holger_gov	
Users			
Groups	Service Operator	amy_ds maria_dev	
	Cluster User	Add User	

There are 4 user personas present in Sandbox:

- maria_dev maria_dev is responsible for preparing and getting insight from data.
 She loves to explore different HDP components like Hive, Pig, HBase, Phoenix, etc.
- 2 raj_ops raj_ops is responsible for infrastructure build and R&D activities like design, install, configure and administration. He serves as a technical expert in the area of system administration for complex operating systems.
- 3 holger_gov holger_gov is primarily for the management of data elements, both the content and metadata. He has a specialist role that incorporates processes, policies, guidelines and responsibilities for administering organizations' entire data in compliance with policy and/or regulatory obligations.
- 4 **amy_ds** A data scientist who uses Hive, Spark and Zeppelin to do exploratory data analysis, data cleanup and transformation as preparation for analysis.

Some notable differences between these users in the Sandbox are mentioned below:

NAME ID(S)	ROLE	SERVICES	NAME ID(S)
Sam Admin	Ambari Admin	Ambari	Sam Admin
Raj (raj_ops)	Hadoop Warehouse	Hive/Tez, Ranger,	Raj (raj_ops)
	Operator	Falcon, Knox, Sqoop,	
		Oozie, Flume,	
		Zookeeper	
Maria (maria_dev)	Spark and SQL	Hive, Zeppelin,	Maria (maria_dev)
	Developer	MapReduce/Tez/Spark,	
		Pig, Solr,	
		HBase/Phoenix, Sqoop,	
		NiFi, Storm, Kafka,	
		Flume	
Amy (amy_ds)	Data Scientist	Spark, Hive, R, Python,	Amy (amy_ds)
		Scala	
Holger (holger_gov)	Data Steward	Atlas	Holger (holger_gov)

OS Level Authorization

NAME ID(S)	AME ID(S) HDFS AUTHORIZATION		RANGER AUTHORIZATION	
Sam Admin	Max Ops	Ambari Admin	Admin access	
Raj (raj_ops)	Access to Hive, Hbase, Atlas, Falcon, Ranger, Knox, Sqoop, Oozie, Flume, Operations	Cluster Administrator	Admin Access	
Maria (maria_dev)	Access to Hive, Hbase, Falcon, Oozie and Spark	Service Operator	Normal User Access	
Amy (amy_ds)	Access to Hive, Spark and Zeppelin	Service Operator	Normal User Access	
Holger (holger_gov)	Access to Atlas	Service Administrator	Normal User Access	

Other Differences

NAME	SANDBOX	VIEW	START/STOP/RE	MODIFY	ADD/DEL	INSTALL	MANAGE	MANA
ID(S)	ROLE	CONFIGURAT	START SERVICE	CONFIGURAT	ETE	COMPONE	USERS/GR	GE
		IONS		IONS	SERVICES	NTS	OUPS	AMBA
								RI
								VIEWS
Sam	Ambari	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Admin	Admin							
Raj	Cluster	Yes	Yes	Yes	Yes	Yes	No	No
(raj_ops)	Administr							
	ator							

Maria	Service	Yes	Yes	No	No	No	No	No
(maria_d	Operator							
ev)								
Amy	Service	Yes	Yes	No	No	No	No	No
(amy_ds)	Operator							
Holger	Service	Yes	Yes	Yes	No	No	No	No
(holger_	Administr							
gov)	ator							