

# Learning the Ropes of the HDP Sandbox

---

## Introduction

This tutorial is aimed for users who do not have much experience in using the Sandbox.

The Sandbox is a straightforward, pre-configured, learning environment that contains the latest developments from Apache Hadoop, specifically the Hortonworks Data Platform (HDP). 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.

Let's begin our Hadoop journey...

## Prerequisites

- Downloaded and deployed [Hortonworks Data Platform \(HDP\) Sandbox](#)

## Outline

- [Environment Setup](#)
- [Terminal Access](#)
- [Welcome Page](#)
- [Explore Ambari](#)
- [Further Reading](#)
- [Appendix A: Reference Sheet](#)
  - [Login Credentials](#)
  - [Sandbox Version](#)
  - [Admin Password Reset](#)
- [Appendix B: Troubleshoot](#)
- [Appendix C: Determine Network Adapter of Your VirtualBox Sandbox](#)

## Environment setup

This is the administrative section to get started with the Hortonworks Sandbox environment. Generally, this will only be done once.

### Determine IP Address of Your Sandbox

Once the Sandbox VM or container is installed, it settles to the host of your environment, the IP address varies depending on your Virtual Machine (VMware, VirtualBox) or container (Docker). Once the sandbox is running, it will tell you the IP address. An example of typical IP addresses for each supported environment:

**Docker:** IP Address = **127.0.0.1**

**VirtualBox:** IP Address = **127.0.0.1**

**VMWare:** IP Address = **192.168.x.x**

If you're using **VirtualBox** or **VMWare**, you can confirm the IP address by waiting for the installation to complete and confirmation screen will tell you the IP address your sandbox resolves to. For example:

```
Hortonworks HDP Sandbox
https://hortonworks.com/products/sandbox

To quickly get started with the Hortonworks Sandbox, follow this tutorial:
https://hortonworks.com/tutorial/hadoop-tutorial-getting-started-with-hdp/

To initiate your Hortonworks Sandbox session, open a browser to this address:

For VirtualBox:
Welcome screen: http://localhost:1000
SSH: http://localhost:4200

For VMWare:
Welcome screen: http://10.0.2.15:1000
SSH: http://10.0.2.15:4200
```

### Map Sandbox IP to Your Desired Hostname in the Hosts File

Mac, Linux and Windows all have a hosts file. This file once configured enables the IP address of the sandbox to be mapped to a hostname that is easier to remember than a number.

**Mac users:**

- `echo '{IP-Address} sandbox-hdp.hortonworks.com sandbox-hdf.hortonworks.com' | sudo tee -a /private/etc/hosts`

#### Linux users:

- `echo '{IP-Address} sandbox-hdp.hortonworks.com sandbox-hdf.hortonworks.com' | sudo tee -a /etc/hosts`

#### Windows users:

- Run Notepad as **administrator**.
- Open **hosts** file located in: `c:\Windows\System32\drivers\etc\hosts`
- Add `{IP-Address} localhost sandbox-hdp.hortonworks.com sandbox-hdf.hortonworks.com`
- Save the file

**IMPORTANT:** Replace `{IP-Address}` with [Sandbox IP Address](#)

## Terminal Access

Refer to [Login Credentials](#) for list of users and passwords. You can also login using **root**, using password **hadoop**, which may require you to change the password - remember it!

If you login using credentials other than **root**, you will be required to use **sudo** before the command. For example: `sudo ambari-server status`.

#### Secure Shell Method:

Open your terminal (mac/linux) or Git Bash (Windows). Type the following command to access the Sandbox through **ssh user@hostname -p port**. For example: `ssh root@sandbox-hdp.hortonworks.com -p 2222`

```
$ssh root@sandbox-hdp.hortonworks.com -p 2222
root@sandbox-hdp.hortonworks.com's password:
Last login: Wed Oct 3 04:34:34 2018 from 172.18.0.3
[root@sandbox-hdp ~]#
```

#### Shell Web Client Method:

The **shell web client** is also known as **Shell-in-a-Box**. It's an easy way to issue shell commands without needing to install additional software. It uses **port 4200**, for example: `sandbox-hdp.hortonworks.com:4200`

#### Send Data Between Sandbox and Local Machine

Using the terminal of your choice, you can transfer files to/from sandbox and local machine.

- Transfer file from local machine to sandbox:
  - `scp -P 2222 <local_directory_file> root@sandbox-hdp.hortonworks.com:<sandbox_directory_file>`
- Transfer file from sandbox to local machine:
  - `scp -P 2222 root@sandbox-hdp.hortonworks.com:<sandbox_directory_file> <local_directory_file>`

Do you notice the difference between the two commands?

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.

## Welcome Page

The Sandbox Welcome Page is also known as the **Splash Page**. It runs on port number **:1080**. To open it, use your host address and append the port number. For example: `http://sandbox-hdp.hortonworks.com:1080/`

It looks like this:

[GET HELP](#)

# SAND BOX HDP



## 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

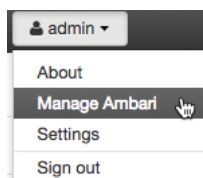
[QUICK LINKS](#)

**Launch Dashboard** opens two browser windows – Ambari interface and beginner tutorial. You should login to Ambari using the username and password based on the tutorial requirements. Most of the tutorials use **raj\_ops** or **maria\_dev**. Refer to [Login Credentials](#) for list of users and passwords.

**Advanced HDP Quick Links** provide quick access to Ambari Services such as Zeppelin, Atlas, Ranger, Shell-in-a-box and others.

## Explore Ambari

- Ambari Dashboard runs on port **:8080**. For example, <http://sandbox-hdp.hortonworks.com:8080>
- Login to as **admin**, refer to [Admin Password Reset](#)
- Select **Manage Ambari**



The following screen will be displayed:

Ambari Admin / Cluster Information

Cluster Name\*  
Sandbox

Cluster Blueprint

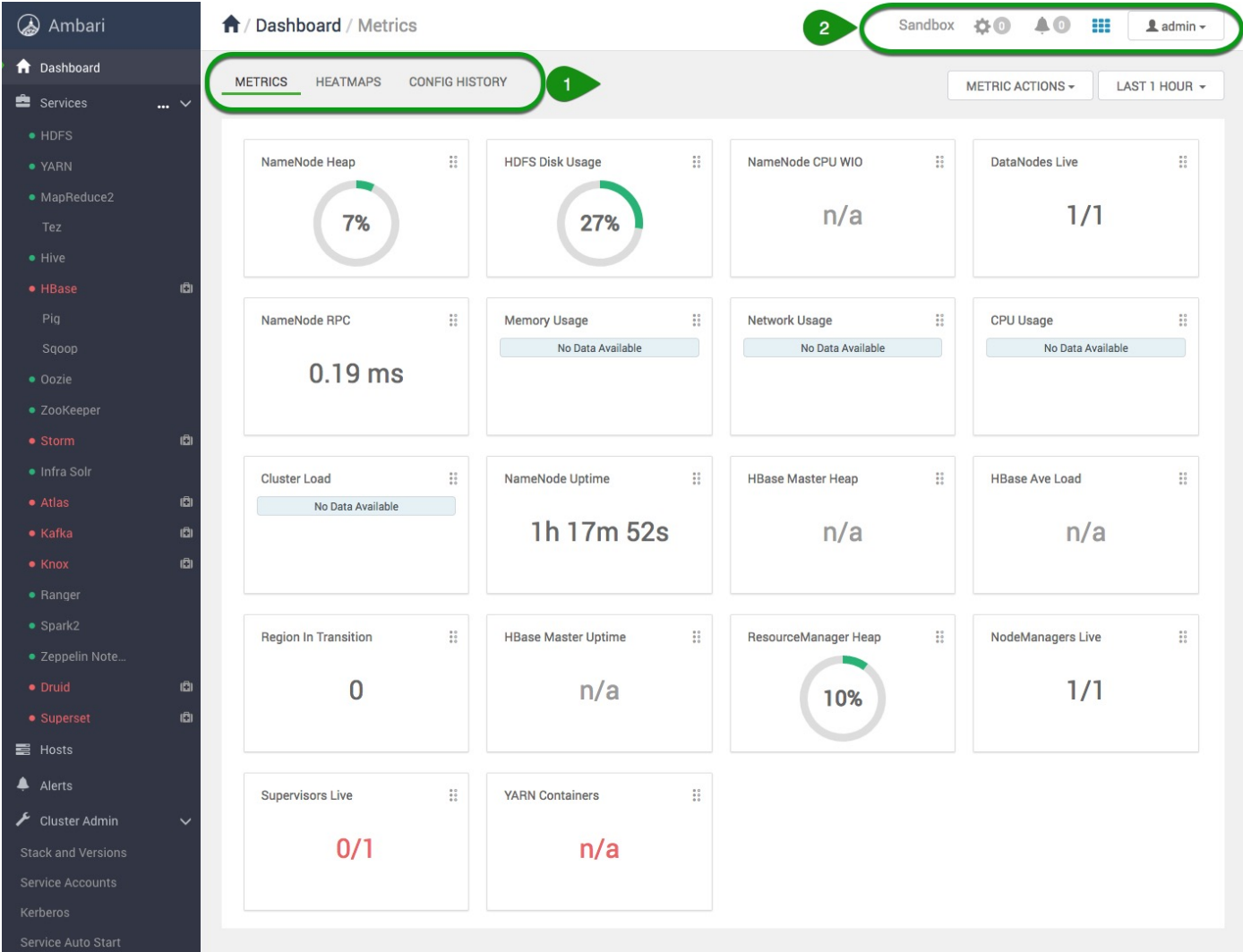
```
{
  "configurations": [
    {
      "zeppelin-log4j-properties": {
        "properties_attributes": {},
        "properties": {
          "log4j_properties_content": "\nlog4j.rootLogger = INFO, dailyfile\nlog4j.appender.stdout = org.apache.log4j.ConsoleAppender\nlog4j.appender.stdout.layout = org.apache.log4j.PatternLayout\nlog4j.appender.stdout.layout.ConversionPattern=%5p [%d] (%t) %F[%M]:%L) - %m%\n\nlog4j.appender.dailyfile.DatePattern=yyyy-MM-dd\nlog4j.appender.dailyfile.Threshold = INFO\nlog4j.appender.dailyfile = org.apache.log4j.DailyRollingFileAppender\nlog4j.appender.dailyfile.File = ${zeppelin.log.file}\nlog4j.appender.dailyfile.layout = org.apache.log4j.PatternLayout\nlog4j.appender.dailyfile.layout.ConversionPattern=%5p [%d] (%t) %F[%M]:%L) - %m%\n"
        }
      },
    },
    {
      "zoo.cfg": {
        "properties_attributes": {},
        "properties": {
          "autopurge.purgeInterval": "24",
          "dataDir": "/hadoop/zookeeper",
          "autopurge.snapRetainCount": "30",

```

1. **"Dashboard"** will take you to the Ambari Dashboard which is the primary UI for Hadoop Operators
2. **"Cluster Management"** allows you to grant permission to Ambari users and groups
3. **"Administration of Users"** allows you to add & remove Ambari users and groups

4. **"Ambari User Views"** list the set of Ambari Users views that are part of the cluster

- Click on **Go to Dashboard** and you should see a similar screen:



Explorer by click on:

1. **Metrics, Heatmaps and Config History**

and then on:

2. **Background Operations, Alerts, Admin and User Views**, icon (represented by 3×3 matrix ) to become familiar with the Ambari resources available to you.

Further Reading

- [Hadoop Tutorial - Getting Started with HDP](#)
- [HDP Documentation](#)
- [Hortonworks Documentation](#)

Appendix A: Reference Sheet

Login Credentials:

User	Password
admin	refer to <a href="#">Admin Password Reset</a>
maria_dev	maria_dev
raj_ops	raj_ops
holger_gov	holger_gov
amy_ds	amy_ds

1. **admin** - System Administrator

2. **maria\_dev** - Responsible for preparing and getting insight from data. She loves to explore different HDP components like Hive, Pig, HBase.

3. **raj\_ops** - Responsible for infrastructure build, research and development activities like design, install, configure and administration. He serves as a technical expert in the area of system administration for complex operating systems.
4. **holger\_gov** - 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.
5. **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
Sam Admin	Ambari Admin	Ambari
Raj (raj_ops)	Hadoop Warehouse Operator	Hive/Tez, Ranger, Falcon, Knox, Sqoop, Oozie, Flume, Zookeeper
Maria (maria_dev)	Spark and SQL Developer	Hive, Zeppelin, MapReduce/Tez/Spark, Pig, Solr, HBase/Phoenix, Sqoop, NiFi, Storm, Kafka, Flume
Amy (amy_ds)	Data Scientist	Spark, Hive, R, Python, Scala
Holger (holger_gov)	Data Steward	Atlas

OS Level Authorization

Name id(s)	HDFS Authorization	Ambari Authorization	Ranger Authorization
Sam Admin	Max Ops	Ambari Admin	Admin access
Raj (raj_ops)	Access to Hive, Hbase, Atlas, Falcon, 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 id(s)	Sandbox Role	View Configurations	Start/Stop/Restart Service	Modify Configurations	Add/delete services	Install Components	Manage Users/Groups	Manage Ambari Views	Atlas UI Access	Sandbox Access
Sam Admin	Ambari Admin	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Full
Raj (raj_ops)	Cluster Administrator	Yes	Yes	Yes	Yes	Yes	No	No	No	Full
Maria (maria_dev)	Service Operator	Yes	Yes	No	No	No	No	No	No	SEI
Amy (amy_ds)	Service Operator	Yes	Yes	No	No	No	No	No	No	SEI
Holger (holger_gov)	Service Administrator	Yes	Yes	Yes	No	No	No	No	Yes	SEI, CRI, DI

Open a port for custom use

Refer to the [Sandbox Guide](#) for ports available for custom use.

In this example we will use the fictional port **1234**, note that this port is not available for custom use.

SSH onto the Sandbox Host

If you are running the VirtualBox VM:

```
# SSH on to VirtualBox Virtual Machine
ssh root@sandbox-hdp.hortonworks.com -p 2200
```

Or if you are using VMWare:

```
# SSH on to VMWare Virtual Machine
ssh root@sandbox-hdp.hortonworks.com -p 22
```

Note: The default password is **hadoop**.

Change directories to `/sandbox/deploy-scripts/assets/`

```
cd /sandbox/deploy-scripts/assets/
```

NOTE: On the docker version of the Sandbox the script is found on `deploy-scripts/assets/generate-proxy-deploy-script.sh`

Under the assets directory you will find a file named `generate-proxy-deploy-script.sh`, edit it:

```
vi generate-proxy-deploy-script.sh
```

search for the `tcpPortsHDP` array and enter the port that you would like to forward:

```
tcpPortsHDP=(
[1234]=1234
[12049]=2049
[2201]=22
[2222]=22
[1100]=1100
[1111]=1111
[12200]=1220
[1988]=1988
[2100]=2100
[2181]=2181
```

to save and exit press `esc` and enter `:x`

to execute your changes re-run the script:

```
cd /sandbox/deploy-scripts
assets/generate-proxy-deploy-script.sh
```

and deploy the reverse proxy with your changes:

```
/sandbox/proxy/proxy-deploy.sh
```

Finally, add the port forward on your virtual environment

Settings -> Network -> Advanced -> Port Forwarding -> Add New

Now restart the virtual machine and enjoy your new port.

## Sandbox Version

When you run into an issue, one of the first things someone will ask is "what sandbox version are you using"? To get this information:

Login using [shell web client](#) and execute: `sandbox-version`. The output should look something like:

```
[root@sandbox-hdp ~]# sandbox-version
== Sandbox Information ==
Platform: hdp-security
Build date: 09-27-2018
Ambari version: 2.7.1.0-169
Hadoop version: Hadoop 3.1.1.3.0.1.0-187
OS: CentOS Linux release 7.5.1804 (Core)
==
[root@sandbox-hdp ~]#
```

Note: refer to [Login Credentials](#)

## Admin Password Reset

Due to possibility of passwords being vulnerable to being hacked, we recommend you change your Ambari admin password to be unique.

1. Open [Shell Web Client](#) (aka Shell-in-a-Box):
2. The login using credentials: **root** / **hadoop**
3. Type the following commands: `ambari-admin-password-reset`

IMPORTANT: The first time you login as **root**, you may be required to change the password - remember it!

## Appendix B: Troubleshoot

- [Hortonworks Community Connection](#) (HCC) is a good resource to find answers to problems you may encounter during your Hadoop journey.
- [hangs / Long running processes](#)

At times you may encounter a job, query or request that just seems to run forever and does not complete. It may be because it's in the **ACCEPTED** state. A good place to begin looking is in the [ResourceManager](#). If you know a job has completed, but the Resource Manager still thinks it's running - kill it!

**memory usage**

**All Applications**

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Lost Nodes
6	0	5	1	5	2.20 GB	2.93 GB	0 B	5	8	0	0

Scheduler Metrics

Capacity Scheduler [MEMORY] Scheduling Resource Type <memory:250, vCores:1> Minimum Alloc

Show: 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalState
application_1510885864807_0006	maria_dev	TempletonControllerJob	MAPREDUCE	default	0	Thu Nov 16 23:28:06 -0800 2017	N/A	ACCEPTED	UNDEFI
application_1510885864807_0005	hive	HIVE-8fd7ea6a-68be-4f37-bb34-9ba9efb0e11	TEZ	default	0	Thu Nov 16 23:25:53 -0800 2017	N/A	RUNNING	UNDEFI
application_1510885864807_0004	hive	HIVE-a619347c-8638-403d-b75f-cb0b38c36ac5	TEZ	default	0	Thu Nov 16 23:24:51 -0800 2017	N/A	RUNNING	UNDEFI
application_1510885864807_0003	hive	HIVE-3abff04d-8df5-425c-a098-39f54cc4a2a	TEZ	default	0	Thu Nov 16 23:24:24 -0800 2017	N/A	RUNNING	UNDEFI
application_1510885864807_0002	hive	HIVE-9a2928f2-a63c-45e9-a160-fdb7bd3e6577	TEZ	default	0	Thu Nov 16 23:16:13 -0800 2017	N/A	RUNNING	UNDEFI
application_1510885864807_0001	hive	HIVE-f8d24f87-6ac1-4f9c-9346-d7cac98a617b	TEZ	default	0	Thu Nov 16 23:15:19 -0800 2017	Thu Nov 16 23:26:33 -0800 2017	FINISHED	SUCCEDEE

Completed long time ago...

Let's kill them by selecting ID and Kill the Application.

**Tools**

- Cluster
- About
- Nodes
- Node Labels
- Applications
- NEW SAVING
- SUBMITTED
- ACCEPTED
- RUNNING
- FINISHED
- FAILED
- KILLED
- Scheduler

**Kill Application**

## Appendix C: Determine Network Adapter of Your VirtualBox Sandbox

Once the Sandbox VM is installed, it attaches to a virtual network. There are 8 different network modes, but the default network your sandbox will attach to is NAT. We will cover relevant networks for our tutorial use cases: NAT and Bridged Adapter.

### Network Address Translation (NAT)

By default, the VM attaches to Network Address Translation (NAT) network mode. The guest's IP address by default translates over to the host's IP address. NAT allows for the guest system to connect to external devices on external networks, but external devices cannot access the guest system. Alternatively, VirtualBox can make selected services on the guest reachable to the outside world by port forwarding. VirtualBox listens to certain ports on the host, then re-sends packets that arrive at those ports to the guest on the same port or different port.

How we are forwarding all incoming traffic from a specific host interface to the guest in our sandbox is by specifying an IP of that host like the following:

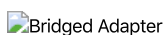
```
VBoxManage modifyvm "Hortonworks Sandbox HDP 3.0.1" --natpf1 "Sandbox Splash Page,tcp,127.0.0.1,1080,,1080"
.
.
.
VBoxManage modifyvm "Hortonworks Sandbox HDP 3.0.1" --natpf1 "Sandbox Host SSH,tcp,127.0.0.1,2122,,22"
```

You can find the set network by opening the VM **settings** and then select the **network** tab.

### Bridged Networking

In this mode, the guest receives direct access to the network, which the host has been connected. The router assigns an IP address to the guest. On that network, instead of there being just the host IP address visible, now the guest IP address is visible too. Thus, external devices, such as MiNiFi running on a Raspberry Pi, are able to connect to the guest via it's IP address.

When would you need this mode? It is needed for Connected Data Architecture(CDA). To configure this mode, first power down your guest vm, click settings, switch to the network tab and change the **attach to** network to be **Bridged Adapter**.



**WARNING:** First make sure your computer is connected to a router, else this feature will not work cause there is no router to assign an IP address to the guest vm.

If you're using **VirtualBox** or **VMWare**, you can confirm the IP address by waiting for the installation to complete and confirmation screen will tell you the IP address your sandbox resolves to. For example:



```
Hortonworks HDP Sandbox
https://hortonworks.com/products/sandbox

To quickly get started with the Hortonworks Sandbox, follow this tutorial:
https://hortonworks.com/tutorial/hadoop-tutorial-getting-started-with-hdp/

To initiate your Hortonworks Sandbox session, open a browser to this address:

For VirtualBox:
Welcome screen: http://localhost:1080 ← Localhost is loopback IP address
SSH: http://localhost:4200

For VMWare:
Welcome screen: http://10.0.2.15:1080
SSH: http://10.0.2.15:4200
↑
Unique IP address of Guest VM
```

Note: Guest VM Welcome Window for NAT Sandbox --->

```
Hortonworks HDP Sandbox
https://hortonworks.com/products/sandbox

To quickly get started with the Hortonworks Sandbox, follow this tutorial:
https://hortonworks.com/tutorial/hadoop-tutorial-getting-started-with-hdp/

To initiate your Hortonworks Sandbox session, open a browser to this address:

For VirtualBox:
Welcome screen: http://localhost:1080 ← Localhost is loopback IP address
SSH: http://localhost:4200

For VMWare:
Welcome screen: http://192.168.1.74:1080
SSH: http://192.168.1.74:4200
↑
Unique IP address of Guest VM
```

Note: Guest VM Welcome Window for BRIDGED Sandbox --->