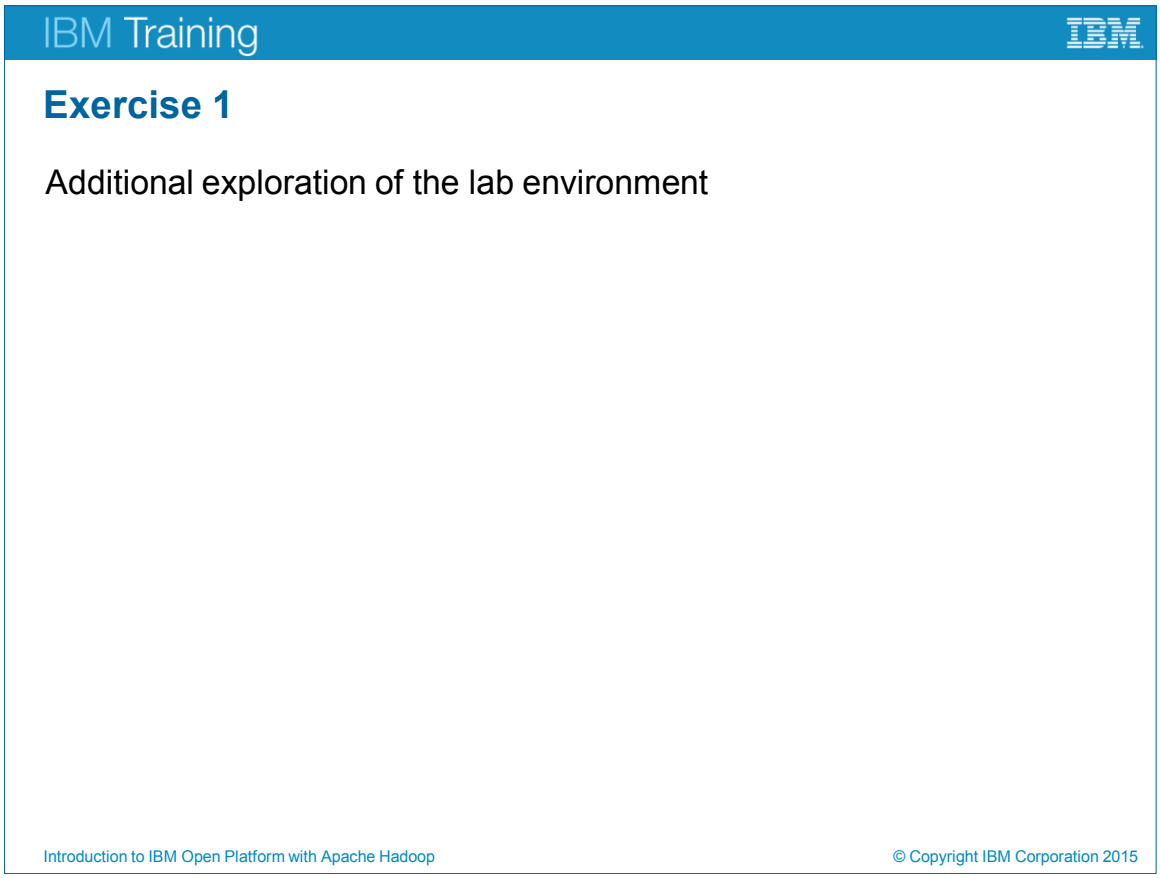


Exercise 1: Additional exploration of the lab environment

Exercise 1: Additional exploration of the lab environment

Purpose:
This is an additional exploratory exercise of the lab environment. You will access your lab environment and use a command line window. You will be exploring the directory structure of the Linux system that you will be using.

VM Hostname: **http://ibmclass.localdomain**
User/Password: **biadmin / biadmin**
root/dalvm3

Task 1. Login to your image and perform some basic setup.

You will need verify that your hostname and IP address are setup correctly, and will make changes if they are needed.

Note: If you powered your VMware image from an earlier session, this verification of hostname and IP address should be repeated.

1. Connect to and login to your VMware image with the user as **biadmin** and the password **biadmin**.

   The desktop of your Linux system is displayed (desktop of the user biadmin).

   To avoid having to constantly login, you can change your screen saver settings.

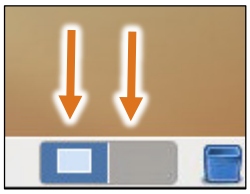
2. At the top left border of your Linux window, choose **System > Preferences > Screensaver**, uncheck the box **Lock screen when screen saver is active**, and then click the **Close** button.

3. To open a terminal window, right-click the desktop and select **Open in Terminal**.

   This terminal window can be resized. You may find it useful to stretch the right side to make it wider, as some commands are long and some responses will be wide; any text that is longer than the default line width will be word-wrapped to the next line.

   You may want to have more than one window open. When you do this, place multiple windows so that they do not overlap completely.

You can also have more than one terminal window, and have the choice of several, simultaneous desktops that you can alternate between:



Here the desktop on the left is currently selected, and the second arrow points to an alternate/extension desktop for the same login session.

4. In your terminal window, type **`ifconfig`**.

   Take note of the IP address (inet addr) for your environment and your Ethernet adapter (eth0, or eth1, etc., or eth4 as in this case).

5. To see the hostname, type the Linux command **`hostname`** in your terminal window.

   The results appear as follows:

```
[biadmin@ibmclass ~]$ hostname
ibmclass.localdomain
```

6. Check your /etc/hosts file by using **cat /etc/hosts** to verify whether the IP address and hostname are recorded correctly.

   The results appear as follows:

```
[biadmin@ibmclass ~]$ cat /etc/hosts
10.0.0.68 ibmclass.localdomain ibmclass
127.0.0.1 localhost.localdomain localhost
```

   This listing shows that the hostname and IP address are correctly set in the /etc/hosts file. If you find that the values in steps 4 and 5 are not the same as what you find in step 6, you will have to edit the /etc/hosts file. Since /etc/hosts is a system file, you will need to edit that file as an administrator (root).

   If your settings in the /etc/host file are correct, proceed to step 8.

7. If you need to edit /etc/hosts, open the file **/etc/host**s as an administrator (root):
   a. Either: **sudo gedit /etc/hosts**
      or:
      ```
      [biadmin@ibmclass ~]$ su -
      Password:
      [root@ibmclass ~]# gedit /etc/hosts
      ```
      For sudo, use your biadmin password. For substitute user (su -), use the root password.
   b. Edit the line that shows the current system to match the information you found in steps 4 and 5.
   c. Save and close.
   d. If you used su, exit from the root login with **Ctrl-D** or **exit**.
8. If you have connectivity problems later in this exercise or in another exercise, perform steps 5 and 6 and edit the /etc/hosts file again, if necessary.

   Repeating these steps will not normally be necessary unless your lab environment is restarted, either deliberately or overnight.

Task 2. Start Ambari and the Hadoop processing environment.

In order to work with your Hadoop cluster environment, you need to start the Ambari server and use Ambari to initialize the cluster.

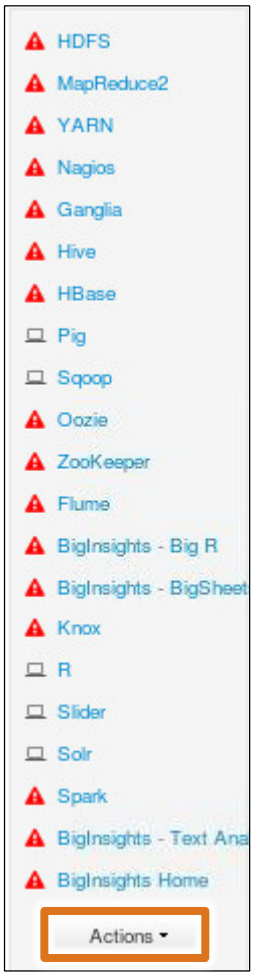
The Ambari server takes some time to go through startup and initialization. While it is starting, you can move on to other activities.

1. Launch **Firefox**, and then if necessary, navigate to the **Ambari** login page, **http://localhost:8080**.

   You are running a single-host Hadoop cluster (pseudo-distributed mode) and thus you can connect to the Ambari server on port 8080.
2. Log in to the **Ambari** console as **admin/admin**.

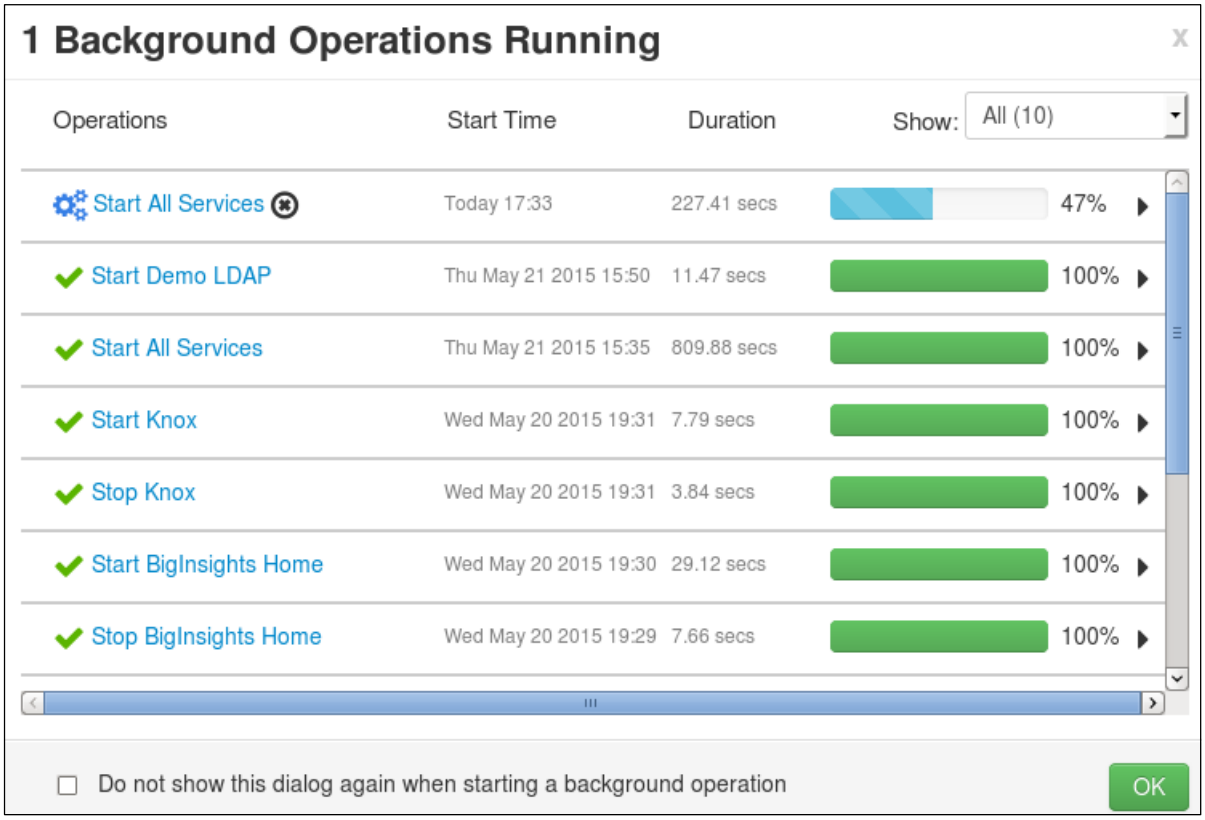
   On the left side of the browser are the statuses of all the services. If they are currently yellow, wait a couple of minutes for them to become red before you start them up.

3. Once all the statuses are red, at the bottom of the left side, click **Actions** and then click **Start All** to start the services.



This will take several minutes to complete.

This will bring up a window showing that Start All Services is running in the background (running tasks are colored blue). Completed tasks are colored green, and failed tasks are colored red.



4. When the services have started successfully, click **OK**, and then close Firefox.

   Starting all services may take as much as 10 minutes.

   You will be revisiting the Ambari Server in its own unit and will look at the various pages in more detail at that time.

   You will now explore your Hadoop cluster environment.

Task 3. Explore the placement of files in the Linux host-system environment.

You will review the placement of files in the Linux host-system environment so that you can become familiar with where the open-source ODP binaries, all configuration files, and the BigInsights software are installed. You need this knowledge to be able to work best with a Hadoop cluster.

This Hadoop cluster is unique however, as it is a single-node cluster. This is called pseudo-distributed mode since it simulates a multi-node cluster.

The Open Data Platform software is found in the /usr/iop directory. This is the Apache open source software that forms the basis for the Open Data Platform Initiative. (IOP = IBM Open Platform).

1. In a new terminal window, type **`cd`** to change to your home directory.
2. Type **`pwd`** to verify that you are in biadmin's home directory.
3. To change the directory to /usr/iop, type **`cd /usr/iop`**.

   The results appear as follows:

```
[biadmin@ibmclass Desktop]$ cd
[biadmin@ibmclass ~]$ pwd
/home/biadmin
[biadmin@ibmclass ~]$ cd /usr/iop
[biadmin@ibmclass iop]$
```

4. To review the files and directories in the /usr/iop directory, type **`ls -l`**.

```
[biadmin@ibmclass iop]$ ls -l
total 8
drwxr-xr-x. 19 root root 4096 Apr 15 13:07 4.0.0.0
drwxr-xr-x.  2 root root 4096 Jun 11 11:32 current
```

The first of these subdirectories (4.0.0.0) is the release level of the ODP software that you are working with; the version in your lab environment may differ if the software has been updated.

5. To display the contents of the 4.0.0.0 directory, type **`ls -l 4*`**.
   The results appear similar to the following:

```
[biadmin@ibmclass iop]$ ls -l 4*
total 68
drwxr-xr-x. 18 root  root 4096 Apr 15 13:07 etc
drwxr-xr-x.  7 root  root 4096 Apr 15 12:45 flume
drwxr-xr-x.  9 root  root 4096 Apr 15 12:51 hadoop
drwxr-xr-x.  6 root  root 4096 Apr 15 12:44 hadoop-hdfs
drwxr-xr-x.  5 root  root 4096 Apr 15 12:43 hadoop-mapreduce
drwxr-xr-x.  6 root  root 4096 Apr 15 12:43 hadoop-yarn
drwxr-xr-x.  8 root  root 4096 Apr 15 12:50 hbase
drwxr-xr-x.  8 root  root 4096 Apr 15 12:51 hive
drwxrwxr-x.  7 hive  hive 4096 Apr 15 12:52 hive-hcatalog
drwxr-xr-x.  8 root  root 4096 Jun 11 11:33 knox
drwxr-xr-x. 11 oozie root 4096 Apr 15 13:19 oozie
drwxr-xr-x.  7 root  root 4096 Apr 15 13:01 pig
drwxr-xr-x.  6 root  root 4096 Apr 15 13:02 slider
drwxr-xr-x.  8 root  root 4096 Apr 15 13:04 solr
drwxr-xr-x.  8 root  root 4096 Apr 15 13:06 spark
drwxr-xr-x.  6 root  root 4096 Apr 15 13:07 sqoop
drwxr-xr-x.  7 root  root 4096 Apr 15 12:43 zookeeper
```

   Take note of the user IDs associated with the various directories. Some have a user name that is same as the software held in the directory; some are owned by root. You will be looking at the standard users in the Apache Ambari unit when you explore details of the Ambari Server and work with the management of your cluster.

6. To view the **current** subdirectory which has a set of links that point back to files and subdirectories in the 4.*.*.* directory, type **`ls -l current`**.

```
[biadmin@ibmclass iop]$ ls -l current
total 0
lrwxrwxrwx 1 root 25 Jun  2 18:02 accumulo-client -> /usr/iop/4.0.0.0/accumulo
lrwxrwxrwx 1 root root 25 Jun  2 18:02 accumulo-gc -> /usr/iop/4.0.0.0/accumulo
lrwxrwxrwx 1 root root 25 Jun  2 18:02 accumulo-master -> /usr/iop/4.0.0.0/accumulo
lrwxrwxrwx 1 root root 25 Jun  2 18:02 accumulo-monitor -> /usr/iop/4.0.0.0/accumulo
lrwxrwxrwx 1 root root 25 Jun  2 18:02 accumulo-tablet -> /usr/iop/4.0.0.0/accumulo
lrwxrwxrwx 1 root root 25 Jun  2 18:02 accumulo-tracer -> /usr/iop/4.0.0.0/accumulo
lrwxrwxrwx 1 root root 23 Jun  2 18:02 falcon-client -> /usr/iop/4.0.0.0/falcon
lrwxrwxrwx 1 root root 23 Jun  2 18:02 falcon-server -> /usr/iop/4.0.0.0/falcon
lrwxrwxrwx 1 root root 22 Jun  2 18:02 flume-server -> /usr/iop/4.0.0.0/flume
lrwxrwxrwx 1 root root 23 Jun  2 18:02 hadoop-client -> /usr/iop/4.0.0.0/hadoop
lrwxrwxrwx 1 root root 28 Jun  2 18:02 hadoop-hdfs-client -> /usr/iop/4.0.0.0/hadoop-
hdfs
lrwxrwxrwx 1 root root 28 Jun  2 18:02 hadoop-hdfs-datanode -> /usr/iop/4.0.0.0/hadoop-
hdfs
lrwxrwxrwx 1 root root 28 Jun  2 18:02 hadoop-hdfs-journalnode ->
/usr/iop/4.0.0.0/hadoop-hdfs
lrwxrwxrwx 1 root root 28 Jun  2 18:02 hadoop-hdfs-namenode -> /usr/iop/4.0.0.0/hadoop-
hdfs
lrwxrwxrwx 1 root root 28 Jun  2 18:02 hadoop-hdfs-nfs3 -> /usr/iop/4.0.0.0/hadoop-hdfs
```

```
lrwxrwxrwx 1 root root 28 Jun  2 18:02 hadoop-hdfs-portmap -> /usr/iop/4.0.0.0/hadoop-
hdfs
lrwxrwxrwx 1 root root 28 Jun  2 18:02 hadoop-hdfs-secondarynamenode ->
/usr/iop/4.0.0.0/hadoop-hdfs
lrwxrwxrwx 1 root root 33 Jun  2 18:02 hadoop-mapreduce-client ->
/usr/iop/4.0.0.0/hadoop-mapreduce
lrwxrwxrwx 1 root root 33 Jun  2 18:02 hadoop-mapreduce-historyserver ->
/usr/iop/4.0.0.0/hadoop-mapreduce
lrwxrwxrwx 1 root root 28 Jun  2 18:02 hadoop-yarn-client -> /usr/iop/4.0.0.0/hadoop-
yarn
lrwxrwxrwx 1 root root 28 Jun  2 18:02 hadoop-yarn-nodemanager ->
/usr/iop/4.0.0.0/hadoop-yarn
lrwxrwxrwx 1 root root 28 Jun  2 18:02 hadoop-yarn-resourcemanager ->
/usr/iop/4.0.0.0/hadoop-yarn
lrwxrwxrwx 1 root root 28 Jun  2 18:02 hadoop-yarn-timelineserver ->
/usr/iop/4.0.0.0/hadoop-yarn
lrwxrwxrwx 1 root root 22 Jun  2 18:02 hbase-client -> /usr/iop/4.0.0.0/hbase
lrwxrwxrwx 1 root root 22 Jun  2 18:02 hbase-master -> /usr/iop/4.0.0.0/hbase
lrwxrwxrwx 1 root root 22 Jun  2 18:02 hbase-regionserver -> /usr/iop/4.0.0.0/hbase
lrwxrwxrwx 1 root root 21 Jun  2 18:02 hive-client -> /usr/iop/4.0.0.0/hive
lrwxrwxrwx 1 root root 21 Jun  2 18:02 hive-metastore -> /usr/iop/4.0.0.0/hive
lrwxrwxrwx 1 root root 21 Jun  2 18:02 hive-server2 -> /usr/iop/4.0.0.0/hive
lrwxrwxrwx 1 root root 30 Jun  2 18:02 hive-webhcat -> /usr/iop/4.0.0.0/hive-hcatalog
lrwxrwxrwx 1 root root 22 Jun  2 18:02 kafka-broker -> /usr/iop/4.0.0.0/kafka
lrwxrwxrwx 1 root root 21 Jun  2 18:02 knox-server -> /usr/iop/4.0.0.0/knox
lrwxrwxrwx 1 root root 23 Jun  2 18:02 mahout-client -> /usr/iop/4.0.0.0/mahout
lrwxrwxrwx 1 root root 22 Jun  2 18:02 oozie-client -> /usr/iop/4.0.0.0/oozie
lrwxrwxrwx 1 root root 22 Jun  2 18:02 oozie-server -> /usr/iop/4.0.0.0/oozie
lrwxrwxrwx 1 root root 24 Jun  2 18:02 phoenix-client -> /usr/iop/4.0.0.0/phoenix
lrwxrwxrwx 1 root root 20 Jun  2 18:02 pig-client -> /usr/iop/4.0.0.0/pig
lrwxrwxrwx 1 root root 29 Jun  2 18:02 ranger-admin -> /usr/iop/4.0.0.0/ranger-admin
lrwxrwxrwx 1 root root 32 Jun  2 18:02 ranger-usersync -> /usr/iop/4.0.0.0/ranger-
usersync
lrwxrwxrwx 1 root root 23 Jun  2 18:02 slider-client -> /usr/iop/4.0.0.0/slider
lrwxrwxrwx 1 root root 22 Jun  2 18:02 spark-client -> /usr/iop/4.0.0.0/spark
lrwxrwxrwx 1 root root 22 Jun  2 18:02 spark-historyserver -> /usr/iop/4.0.0.0/spark
lrwxrwxrwx 1 root root 22 Jun  2 18:02 spark-thriftserver -> /usr/iop/4.0.0.0/spark
lrwxrwxrwx 1 root root 22 Jun  2 18:02 sqoop-client -> /usr/iop/4.0.0.0/sqoop
lrwxrwxrwx 1 root root 22 Jun  2 18:02 sqoop-server -> /usr/iop/4.0.0.0/sqoop
lrwxrwxrwx 1 root root 22 Jun  2 18:02 storm-client -> /usr/iop/4.0.0.0/storm
lrwxrwxrwx 1 root root 22 Jun  2 18:02 storm-nimbus -> /usr/iop/4.0.0.0/storm
lrwxrwxrwx 1 root root 36 Jun  2 18:02 storm-slider-client -> /usr/iop/4.0.0.0/storm-
slider-client
lrwxrwxrwx 1 root root 22 Jun  2 18:02 storm-supervisor -> /usr/iop/4.0.0.0/storm
lrwxrwxrwx 1 root root 20 Jun  2 18:02 tez-client -> /usr/iop/4.0.0.0/tez
lrwxrwxrwx 1 root root 26 Jun  2 18:02 zookeeper-client -> /usr/iop/4.0.0.0/zookeeper
lrwxrwxrwx 1 root root 26 Jun  2 18:02 zookeeper-server -> /usr/iop/4.0.0.0/zookeeper
[biadmin@ibmclass iop]$
```

There are sometimes multiple links back to the one directory.

The software that will be executed at this time (current) is determined by these links. Since you can have multiple versions of software installed (4.0.0,0, 4.1.0.0, etc.), it possible to install the software for an upcoming upgrade and then later have Ambari install a rolling upgrade that will systematically change the links as appropriate to files in the new, upgraded software directories.

7. Close all open windows.

Results:
You explored the lab environment, and the command line window. You also explored the directory structure of the Linux system.