

# Hadoop Basics with InfoSphere BigInsights

*Lesson 2: Hadoop architecture*



## **An IBM Proof of Technology**

---

Catalog Number

---

## Contents

Lab 1Exploring Hadoop Distributed File System.....	4
1.1Getting Started.....	5
1.2Exploring Hadoop Distributed File System (Terminal).....	9
1.2.1Using the command line Interface.....	9
1.3Exploring Hadoop Distributed File System (Web Console).....	15
1.3.1Using the Web Console.....	15
1.3.2Working with the Welcome page.....	16
1.3.3Administering BigInsights.....	18
1.3.4Inspecting the status of your cluster.....	18
1.3.5Starting and stopping a component.....	19
1.3.6Working with Files.....	20
1.4Summary.....	24

---

## Lab 1 Exploring Hadoop Distributed File System

The overwhelming trend towards digital services, combined with cheap storage, has generated massive amounts of data that enterprises need to effectively gather, process, and analyze. Data analysis techniques from the data warehouse and high-performance computing communities are invaluable for many enterprises, however often times their cost or complexity of scale-up discourages the accumulation of data without an immediate need. As valuable knowledge may nevertheless be buried in this data, related scaled-up technologies have been developed. Examples include Google's MapReduce, and the open-source implementation, Apache Hadoop.

Hadoop is an open-source project administered by the Apache Software Foundation. Hadoop's contributors work for some of the world's biggest technology companies. That diverse, motivated community has produced a collaborative platform for consolidating, combining and understanding data.

Technically, Hadoop consists of two key services: data storage using the Hadoop Distributed File System (HDFS) and large-scale parallel data processing using a technique called MapReduce.

After completing this hands-on lab, you will be able to:

- Use Hadoop commands to explore HDFS on the Hadoop system
- Use Web Console to explore HDFS on the Hadoop system

Allow 45 minutes to 1 hour complete this section of lab.

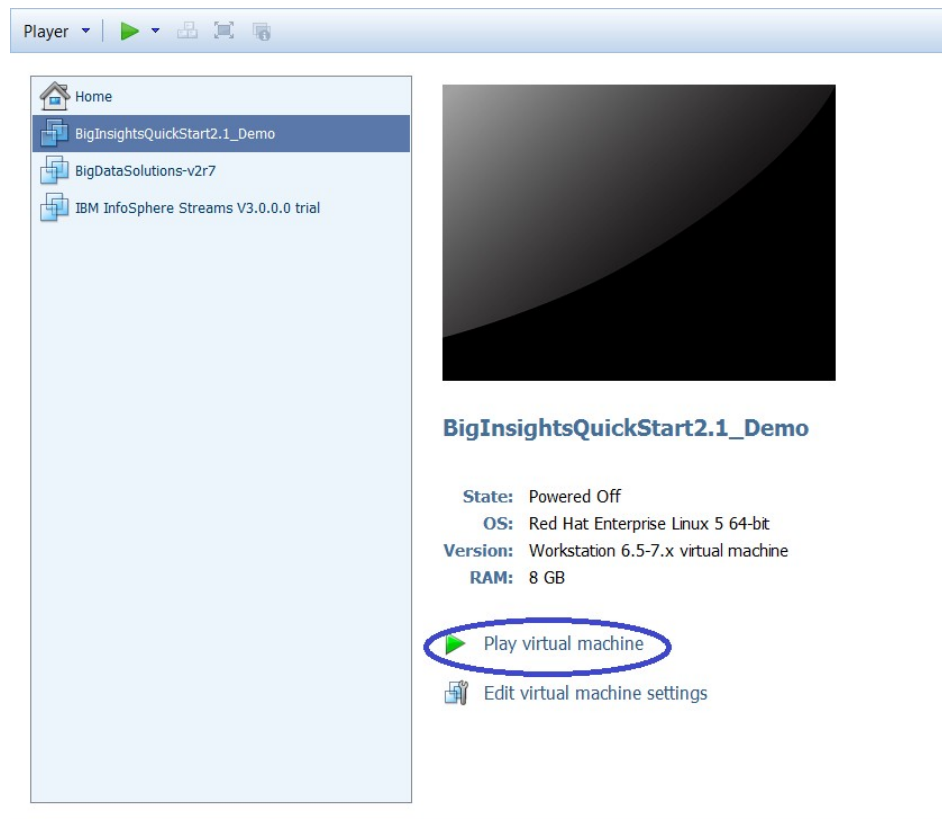
This version of the lab was designed using the InfoSphere BigInsights 2.1 Quick Start Edition. Throughout this lab you will be using the following account login information:

	<b>Username</b>	<b>Password</b>
VM image setup screen	root	password
Linux	biadmin	biadmin

## 1.1 Getting Started

To prepare for the contents of this lab, you must go through the process of getting all of the Hadoop components started

1. Start the VMware image by clicking the Play virtual machine button in the VMware Player if it is not already on.



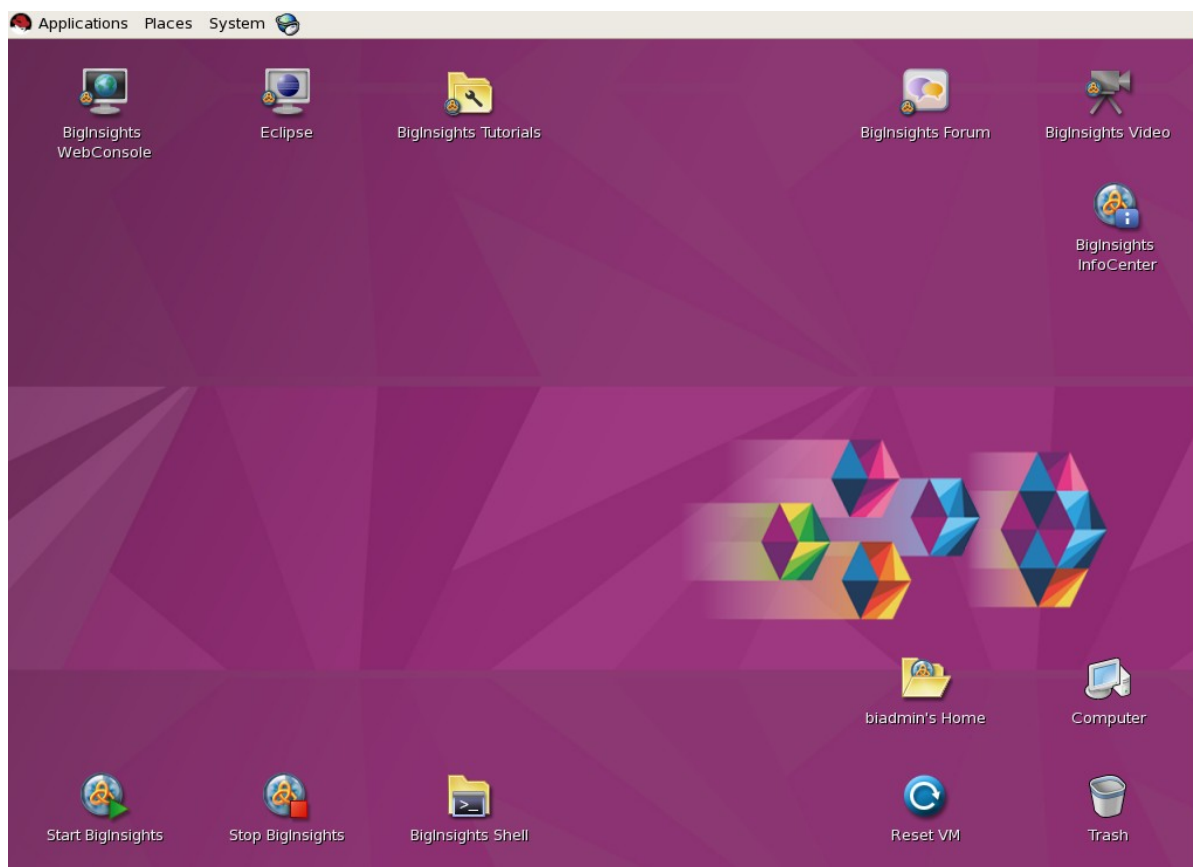
2. Log in to the VMware virtual machine using the following credentials.

User: biadmin

Password: biadmin



\_\_3. After you log in, your screen should look similar to the one below.



Before we can start working with Hadoop Distributed File system, we must first start all the Biginsights components. There are two ways of doing this, through terminal and through simply double-clicking an icon. Both of these methods will be shown in the following steps.

\_\_4. Now open the terminal by double clicking the *BigInsights Shell* icon.



\_\_5. Click on the *Terminal* icon



- \_\_\_6. Once the terminal has been opened change to the \$BIGINSIGHTS\_HOME/bin directory (which by default is /opt/ibm/biginsights)

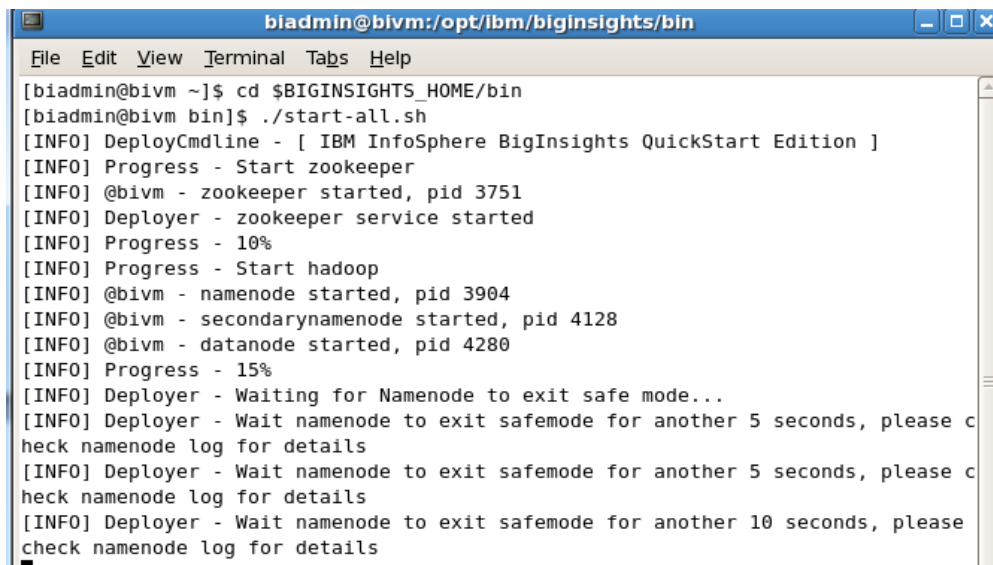
```
cd $BIGINSIGHTS_HOME/bin
```

or

```
cd /opt/ibm/biginsights/bin
```

- \_\_\_7. Start the Hadoop components (daemons) on the BigInsights server. You can practice starting all components with these commands. Please note that they will take a few minutes to run.

```
./start-all.sh
```



```

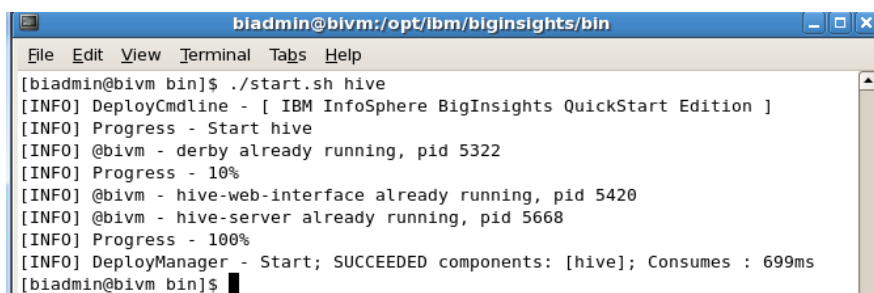
biadmin@bivm:/opt/ibm/biginsights/bin
File Edit View Terminal Tabs Help
[biadmin@bivm ~]$ cd $BIGINSIGHTS_HOME/bin
[biadmin@bivm bin]$ ./start-all.sh
[INFO] DeployCmdline - [ IBM InfoSphere BigInsights QuickStart Edition ]
[INFO] Progress - Start zookeeper
[INFO] @bivm - zookeeper started, pid 3751
[INFO] Deployer - zookeeper service started
[INFO] Progress - 10%
[INFO] Progress - Start hadoop
[INFO] @bivm - namenode started, pid 3904
[INFO] @bivm - secondarynamenode started, pid 4128
[INFO] @bivm - datanode started, pid 4280
[INFO] Progress - 15%
[INFO] Deployer - Waiting for Namenode to exit safe mode...
[INFO] Deployer - Wait namenode to exit safemode for another 5 seconds, please c
check namenode log for details
[INFO] Deployer - Wait namenode to exit safemode for another 5 seconds, please c
check namenode log for details
[INFO] Deployer - Wait namenode to exit safemode for another 10 seconds, please
check namenode log for details

```

- \_\_\_8. Sometimes certain hadoop components may fail to start. You can start and stop the failed components one at a time by using **start.sh** and **stop.sh** respectively. For example to start and stop Hive use:

```
./start.sh hive
```

```
./stop.sh hive
```



```

biadmin@bivm:/opt/ibm/biginsights/bin
File Edit View Terminal Tabs Help
[biadmin@bivm bin]$ ./start.sh hive
[INFO] DeployCmdline - [ IBM InfoSphere BigInsights QuickStart Edition ]
[INFO] Progress - Start hive
[INFO] @bivm - derby already running, pid 5322
[INFO] Progress - 10%
[INFO] @bivm - hive-web-interface already running, pid 5420
[INFO] @bivm - hive-server already running, pid 5668
[INFO] Progress - 100%
[INFO] DeployManager - Start; SUCCEEDED components: [hive]; Consumes : 699ms
[biadmin@bivm bin]$

```

Notice that since Hive did not initially fail, the terminal is telling us that Hive is already running.

- \_\_9. Once all components have started successfully you may move on.

```
[INFO] Deployer - https service started
[INFO] Progress - 100%
[INFO] DeployManager - Start; SUCCEEDED components: [zookeeper, hadoop, derby, h
ive, hbase, bigsql, oozie, orchestrator, console, httpfs]; Consumes : 177411ms
[biadmin@bivm bin]$ clear
```

- \_\_10. If you would like to stop all components execute the command below. However, for this lab please leave all components started.

```
./stop-all.sh
```

Next, let us look at how you would start all the components by double-clicking an icon.

- \_\_11. Double-clicking on the *Start BigInsights* icon would execute a script that does the above mentioned steps. Once all components are started the terminal exits and you are set. Simple.



- \_\_12. We can stop the components in a similar manner, by double-clicking on the *Stop Biginsights* icon. (To the right of *Start BigInsights* icon)



Now that are components are started you may move on to the next section.



## 1.2 Exploring Hadoop Distributed File System (Terminal)

Hadoop Distributed File System (HDFS) allows user data to be organized in the form of files and directories. It provides a command line interface called FS shell that lets a user interact with the data in HDFS accessible to Hadoop MapReduce programs.

There are two methods to interact with HDFS:

1. You can use the command-line approach and invoke the FileSystem (fs) shell using the format:  
*hadoop fs <args>*
2. You can also manipulate HDFS using the BigInsights Web Console.

We will be using both methods in this lab

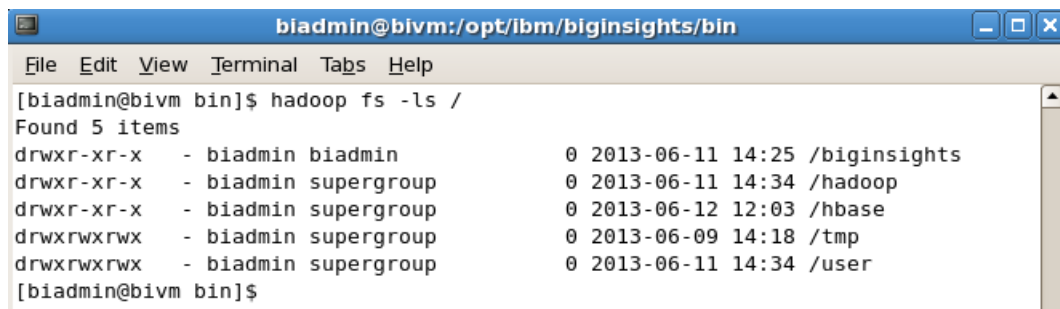
### 1.2.1 Using the command line Interface

We will start with the **hadoop fs -ls** command, which returns the list of files and directories with permission information.

Ensure the Hadoop components are all started, and from the same terminal window as before (and logged on as biadmin), follow these instructions

- \_\_1. List the contents of the root directory.

```
hadoop fs -ls /
```



```
biadmin@bivm:/opt/ibm/biginsights/bin
File Edit View Terminal Tabs Help
[biadmin@bivm bin]$ hadoop fs -ls /
Found 5 items
drwxr-xr-x - biadmin biadmin          0 2013-06-11 14:25 /biginsights
drwxr-xr-x - biadmin supergroup       0 2013-06-11 14:34 /hadoop
drwxr-xr-x - biadmin supergroup       0 2013-06-12 12:03 /hbase
drwxrwxrwx - biadmin supergroup       0 2013-06-09 14:18 /tmp
drwxrwxrwx - biadmin supergroup       0 2013-06-11 14:34 /user
[biadmin@bivm bin]$
```

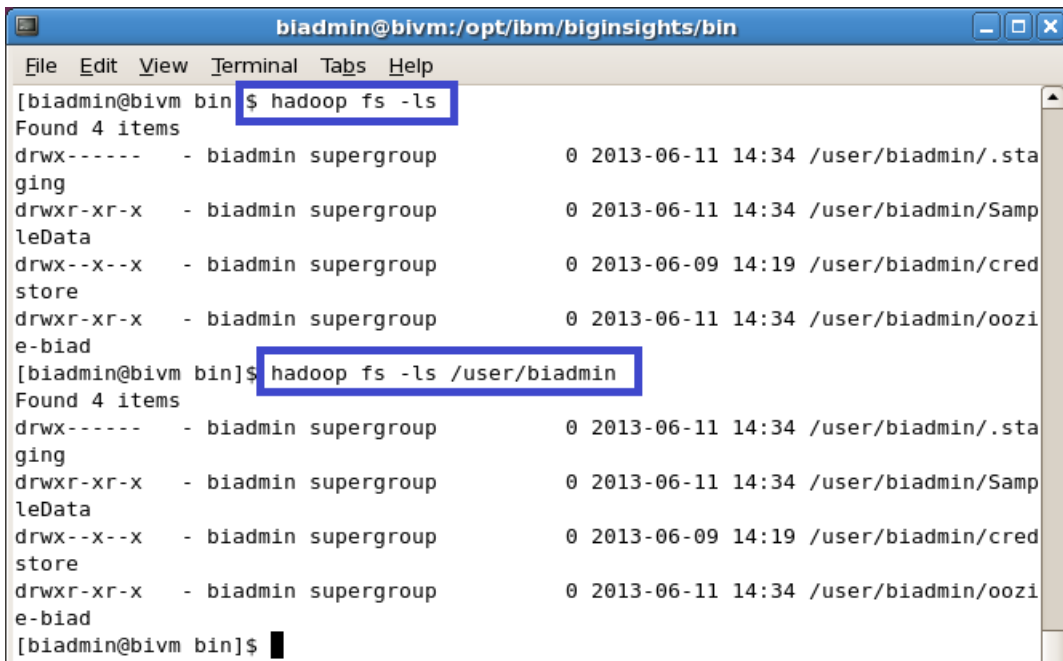
- \_\_2. To list the contents of the /user/biadmin directory, execute:

```
hadoop fs -ls
```

**or**

```
hadoop fs -ls /user/biadmin
```

Note that in the first command there was no director referenced, but it is equivalent to the second command where /user/biadmin is explicitly specified. Each user will get its own home directory under /user. For example, in the case of user biadmin, the home directory is /user/biadmin. Any command where there is no explicit directory specified will be relative to the user's home directory. User space in the native file system (Linux) is generally found under /home/biadmin or /usr/biadmin, but in HDFS user space is /user/biadmin (spelled as "user" rather than "usr").



```

biadmin@bivm:/opt/ibm/biginsights/bin
File Edit View Terminal Tabs Help
[biadmin@bivm bin]$ hadoop fs -ls
Found 4 items
drwx----- - biadmin supergroup          0 2013-06-11 14:34 /user/biadmin/.sta
ging
drwxr-xr-x - biadmin supergroup          0 2013-06-11 14:34 /user/biadmin/Samp
leData
drwx--x--x - biadmin supergroup          0 2013-06-09 14:19 /user/biadmin/cred
store
drwxr-xr-x - biadmin supergroup          0 2013-06-11 14:34 /user/biadmin/oozi
e-biad
[biadmin@bivm bin]$ hadoop fs -ls /user/biadmin
Found 4 items
drwx----- - biadmin supergroup          0 2013-06-11 14:34 /user/biadmin/.sta
ging
drwxr-xr-x - biadmin supergroup          0 2013-06-11 14:34 /user/biadmin/Samp
leData
drwx--x--x - biadmin supergroup          0 2013-06-09 14:19 /user/biadmin/cred
store
drwxr-xr-x - biadmin supergroup          0 2013-06-11 14:34 /user/biadmin/oozi
e-biad
[biadmin@bivm bin]$

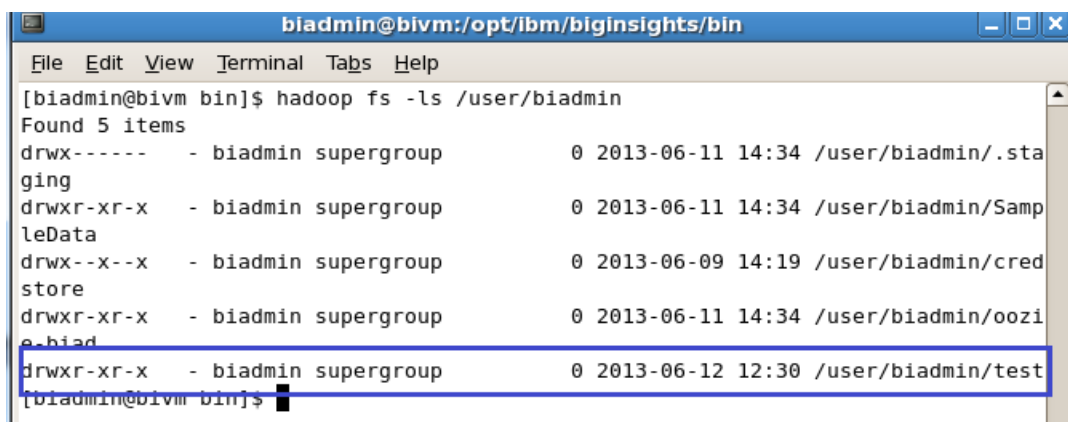
```

\_\_3. To create the directory **test** you can issue the following command:

```
hadoop fs -mkdir test
```

\_\_4. Issue the ls command again to see the subdirectory **test**:

```
hadoop fs -ls /user/biadmin
```



```

biadmin@bivm:/opt/ibm/biginsights/bin
File Edit View Terminal Tabs Help
[biadmin@bivm bin]$ hadoop fs -ls /user/biadmin
Found 5 items
drwx----- - biadmin supergroup          0 2013-06-11 14:34 /user/biadmin/.sta
ging
drwxr-xr-x - biadmin supergroup          0 2013-06-11 14:34 /user/biadmin/Samp
leData
drwx--x--x - biadmin supergroup          0 2013-06-09 14:19 /user/biadmin/cred
store
drwxr-xr-x - biadmin supergroup          0 2013-06-11 14:34 /user/biadmin/oozi
e-biad
drwxr-xr-x - biadmin supergroup          0 2013-06-12 12:30 /user/biadmin/test
[biadmin@bivm bin]$

```

The result of `ls` here is similar to that found with Linux, except for the second column (in this case either “1” or “-”). The “1” indicates the replication factor (generally “1” for pseudo-distributed clusters and “3” for distributed clusters); directory information is kept in the namenode and thus not subject to replication (hence “-”).

To use HDFS commands recursively generally you add an “r” to the HDFS command.

- \_\_5. For example, to do a recursive listing we'll use the `-lsr` command rather than just `-ls`, like the example below.

```
hadoop fs -ls /user
```

```
hadoop fs -lsr /user
```

```

biadmin@bivm:/opt/ibm/biginsights/bin
File Edit View Terminal Tabs Help
[biadmin@bivm bin]$ hadoop fs -ls /user/biadmin
Found 5 items
drwx----- - biadmin supergroup          0 2013-06-11 14:34 /user/biadmin/.sta
ging
drwxr-xr-x - biadmin supergroup          0 2013-06-11 14:34 /user/biadmin/Samp
leData
drwx--x--x - biadmin supergroup          0 2013-06-09 14:19 /user/biadmin/cred
store
drwxr-xr-x - biadmin supergroup          0 2013-06-11 14:34 /user/biadmin/oozi
e-biadm
drwxr-xr-x - biadmin supergroup          0 2013-06-12 12:30 /user/biadmin/test
[biadmin@bivm bin]$ hadoop fs -lsr /user/biadmin
drwx----- - biadmin supergroup          0 2013-06-11 14:34 /user/biadmin/.sta
ging
drwxr-xr-x - biadmin supergroup          0 2013-06-11 14:34 /user/biadmin/Samp
leData
-rw-r--r-- 1 biadmin supergroup      1116423 2013-06-11 13:49 /user/biadmin/Samp
leData/WatsonNewsBlogsData.json
-rw-r--r-- 1 biadmin supergroup      219503 2013-06-11 13:56 /user/biadmin/Samp
leData/WatsonSamle.json

```

- \_\_6. You can pipe (using the `|` character) any HDFS command to be used with the Linux shell. For example, you can easily use `grep` with HDFS by doing the following.

```
hadoop fs -mkdir /user/biadmin/test2
```

```
hadoop fs -ls /user/biadmin | grep test
```

```

biadmin@bivm:/opt/ibm/biginsights/bin
File Edit View Terminal Tabs Help
[biadmin@bivm bin]$ hadoop fs -mkdir /user/biadmin/test2
[biadmin@bivm bin]$ hadoop fs -ls /user/biadmin | grep test
drwxr-xr-x - biadmin supergroup          0 2013-06-12 12:30 /user/biadmin/test
drwxr-xr-x - biadmin supergroup          0 2013-06-12 12:40 /user/biadmin/test
2
[biadmin@bivm bin]$

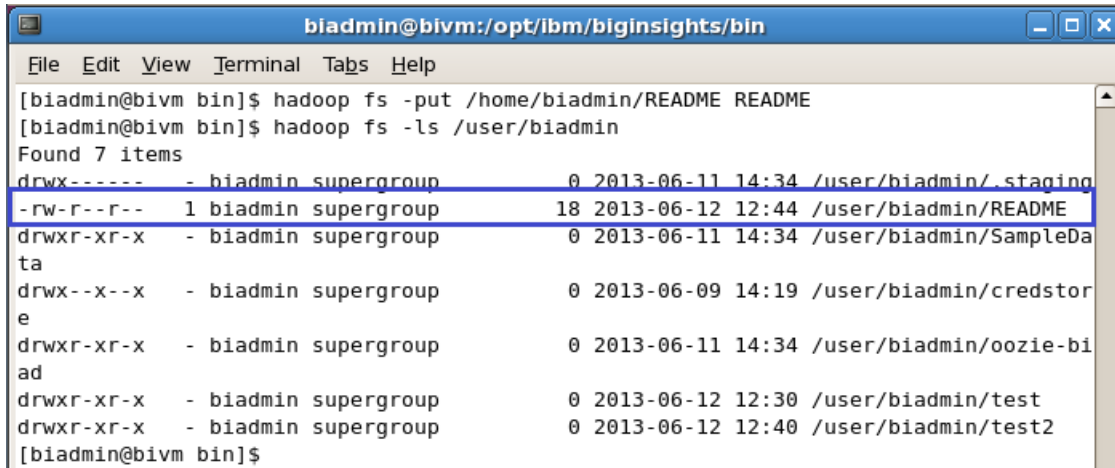
```

As you can see the `grep` command only returned the lines which had “test” in them (thus removing the “Found x items” line and the other directories from the listing).

- \_\_7. To move files between your regular Linux file system and HDFS you can use the put and get commands. For example, move the text file README to the hadoop file system:

```
hadoop fs -put /home/biadmin/README README
```

```
hadoop fs -ls /user/biadmin
```



```

biadmin@bivm:/opt/ibm/biginsights/bin
File Edit View Terminal Tabs Help
[biadmin@bivm bin]$ hadoop fs -put /home/biadmin/README README
[biadmin@bivm bin]$ hadoop fs -ls /user/biadmin
Found 7 items
drwx----- - biadmin supergroup          0 2013-06-11 14:34 /user/biadmin/.staging
-rw-r--r--  1 biadmin supergroup          18 2013-06-12 12:44 /user/biadmin/README
drwxr-xr-x - biadmin supergroup          0 2013-06-11 14:34 /user/biadmin/SampleData
drwx--x--x - biadmin supergroup          0 2013-06-09 14:19 /user/biadmin/credstore
drwxr-xr-x - biadmin supergroup          0 2013-06-11 14:34 /user/biadmin/oozie-bi
drwxr-xr-x - biadmin supergroup          0 2013-06-12 12:30 /user/biadmin/test
drwxr-xr-x - biadmin supergroup          0 2013-06-12 12:40 /user/biadmin/test2
[biadmin@bivm bin]$

```

You should now see a new file called /user/biadmin/README listed as shown above.

- \_\_8. In order to view the contents of this file use the -cat command as follows:

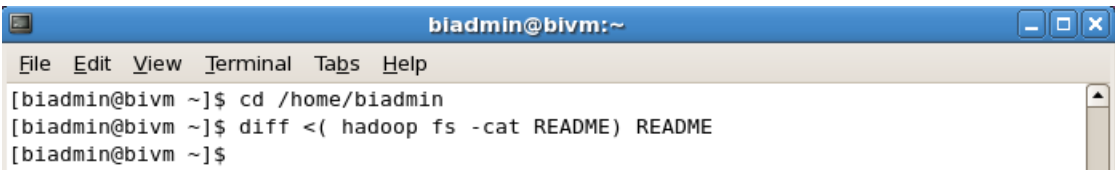
```
hadoop fs -cat README
```

You should see the output of the README file (that is stored in HDFS). We can also use the linux diff command to see if the file we put in HDFS is actually the same as the original on the local filesystem.

- \_\_9. Execute the commands below to use the diff command.

```
cd /home/biadmin/
```

```
diff <( hadoop fs -cat README ) README
```



```

biadmin@bivm:~
File Edit View Terminal Tabs Help
[biadmin@bivm ~]$ cd /home/biadmin
[biadmin@bivm ~]$ diff <( hadoop fs -cat README ) README
[biadmin@bivm ~]$

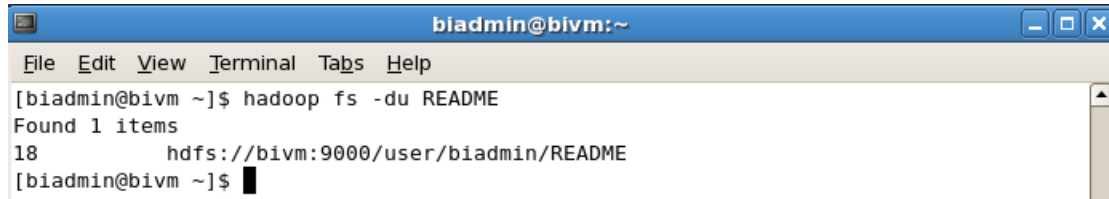
```

Since the diff command produces no output we know that the files are the same (the diff command prints all the lines in the files that differ).

To find the size of files you need to use the -du or -dus commands. Keep in mind that these commands return the file size in bytes.

- \_\_10. To find the size of the README file use the following command.

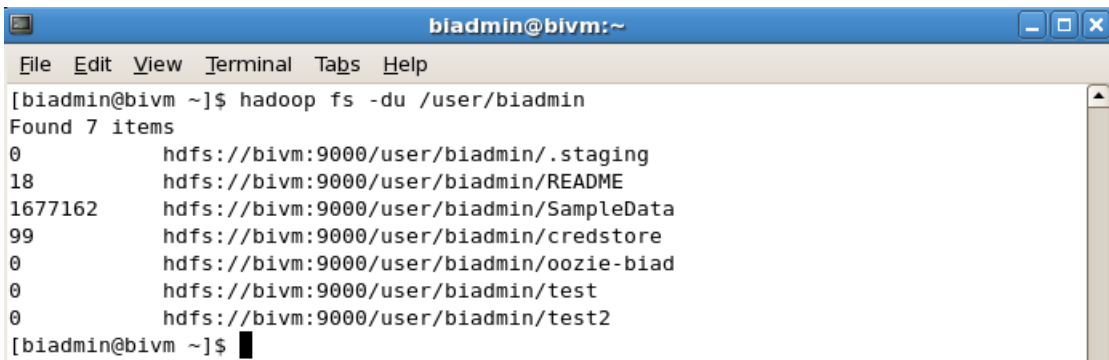
```
hadoop fs -du README
```



```
biadmin@bivm:~  
File Edit View Terminal Tabs Help  
[biadmin@bivm ~]$ hadoop fs -du README  
Found 1 items  
18      hdfs://bivm:9000/user/biadmin/README  
[biadmin@bivm ~]$
```

- \_\_11. To find the size of all files individually in the /user/biadmin directory use the following command:

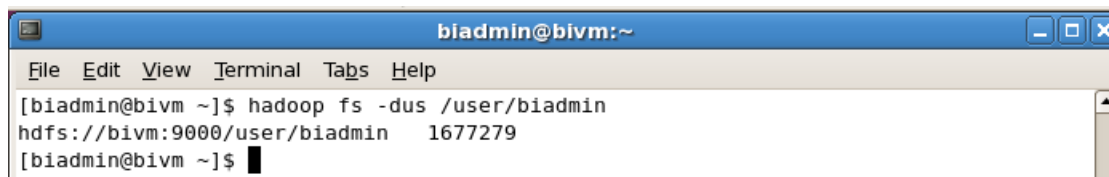
```
hadoop fs -du /user/biadmin
```



```
biadmin@bivm:~  
File Edit View Terminal Tabs Help  
[biadmin@bivm ~]$ hadoop fs -du /user/biadmin  
Found 7 items  
0      hdfs://bivm:9000/user/biadmin/.staging  
18     hdfs://bivm:9000/user/biadmin/README  
1677162 hdfs://bivm:9000/user/biadmin/SampleData  
99     hdfs://bivm:9000/user/biadmin/credstore  
0      hdfs://bivm:9000/user/biadmin/oozie-biad  
0      hdfs://bivm:9000/user/biadmin/test  
0      hdfs://bivm:9000/user/biadmin/test2  
[biadmin@bivm ~]$
```

- \_\_12. To find the size of all files in total of the /user/biadmin directory use the following command.

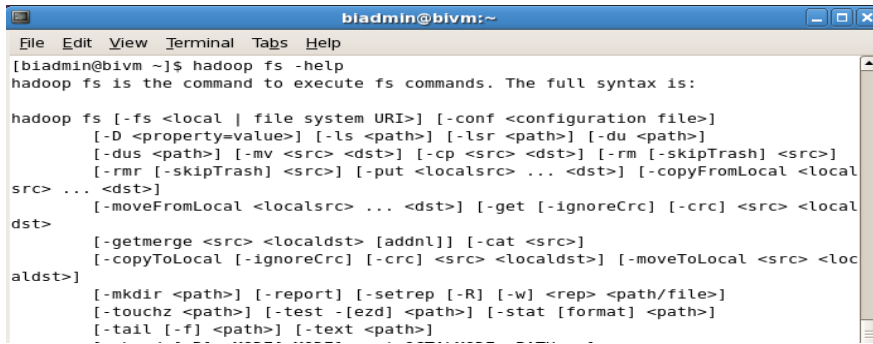
```
hadoop fs -dus /user/biadmin
```



```
biadmin@bivm:~  
File Edit View Terminal Tabs Help  
[biadmin@bivm ~]$ hadoop fs -dus /user/biadmin  
hdfs://bivm:9000/user/biadmin 1677279  
[biadmin@bivm ~]$
```

- \_\_13. If you would like to get more information about hadoop fs commands, invoke -help as follows.

```
hadoop fs -help
```



```

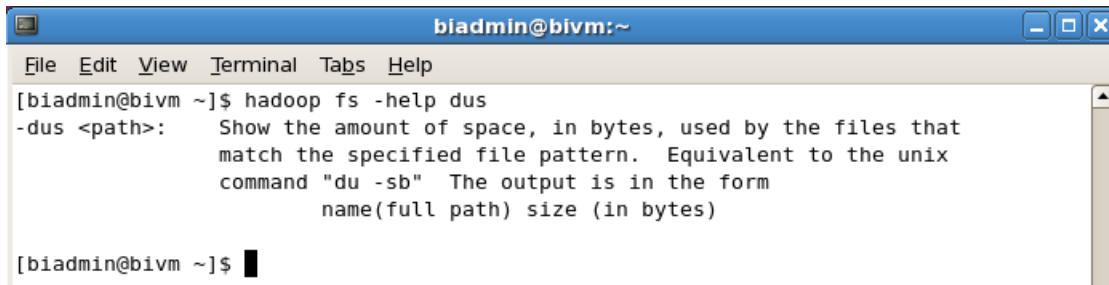
biadmin@bivm:~
File Edit View Terminal Tabs Help
[biadmin@bivm ~]$ hadoop fs -help
hadoop fs is the command to execute fs commands. The full syntax is:

hadoop fs [-fs <local | file system URI>] [-conf <configuration file>]
  [-D <property=value>] [-ls <path>] [-lsr <path>] [-du <path>]
  [-dus <path>] [-mv <src> <dst>] [-cp <src> <dst>] [-rm [-skipTrash] <src>]
  [-rmr [-skipTrash] <src>] [-put <localsrc> ... <dst>] [-copyFromLocal <local
src> ... <dst>]
  [-moveFromLocal <localsrc> ... <dst>] [-get [-ignoreCrc] [-crc] <src> <local
dst>]
  [-getmerge <src> <localdst> [addnl]] [-cat <src>]
  [-copyToLocal [-ignoreCrc] [-crc] <src> <localdst>] [-moveToLocal <src> <loc
aldst>]
  [-mkdir <path>] [-report] [-setrep [-R] [-w] <rep> <path/file>]
  [-touchz <path>] [-test [-ezd] <path>] [-stat [format] <path>]
  [-tail [-f] <path>] [-text <path>]

```

- \_\_\_14. For specific help on a command, add the command name after help. For example, to get help on the `dus` command you'd do the following.

```
hadoop fs -help dus
```



```

biadmin@bivm:~
File Edit View Terminal Tabs Help
[biadmin@bivm ~]$ hadoop fs -help dus
-dus <path>:      Show the amount of space, in bytes, used by the files that
                  match the specified file pattern. Equivalent to the unix
                  command "du -sb" The output is in the form
                  name(full path) size (in bytes)

[biadmin@bivm ~]$ █

```

We are now done with the terminal section, you may close the terminal.

## 1.3 Exploring Hadoop Distributed File System (Web Console)

The first step to accessing the BigInsights Web Console is to launch all of the BigInsights processes (Hadoop, Hive, Oozie, Map/Reduce etc.) They should have been started at the beginning of this lab.

### 1.3.1 Using the Web Console

- \_\_1. Start the Web Console by double-clicking on the *BigInsights WebConsole icon*.



- \_\_2. Verify that your Web console appears similar to this, and note each section:  
**Tasks:** quick access to popular BigInsights tasks,  
**Quick Links:** Links to internal and external quick links and downloads to enhance your environment, and  
**Learn More:** Online resources available to learn more about BigInsights



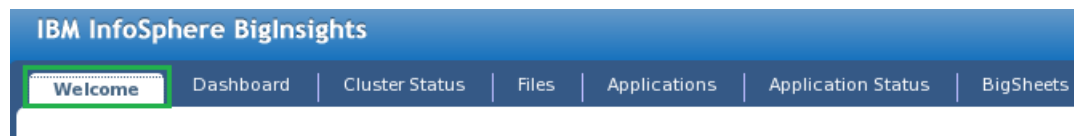
### 1.3.2 Working with the Welcome page

This section introduces you to the Web console's main page displayed through the Welcome tab. The Welcome page features links to common tasks, many of which can also be launched from other areas of the console. In addition, the Welcome page includes links to popular external resources, such as the BigInsights InfoCenter (product documentation) and community forum. You'll explore several aspects of this page.

- \_\_\_3. In the Welcome Tab, the Tasks pane allows you to quickly access common tasks. Select the **View, start or stop a service task**. If necessary scroll down.



- \_\_\_4. This takes you to the **Cluster Status** tab. Here, you can stop and start Hadoop services, as well as gain additional information as shown in the next section
- \_\_\_5. Click on the **Welcome** tab to return back to the main page.

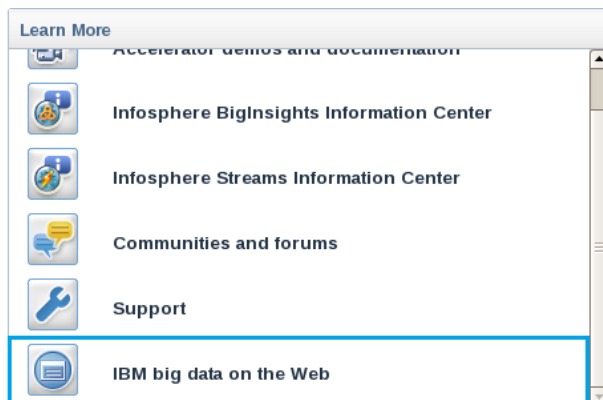




- \_\_\_6. Inspect the **Quick Links** pane at top right and use its vertical scroll bar (if necessary) to become familiar with the various resources accessible through this pane. The first several links simply activate different tabs in the Web console, while subsequent links enable you to perform set-up functions, such as adding BigInsights plug-ins to your Eclipse development environment.



- \_\_\_7. Inspect the **Learn More** pane at lower right. Links in this area access external Web resources that you may find useful, such as the Accelerator demos and documentation, BigInsights InfoCenter, a public discussion forum, IBM support, and IBM's BigInsights product site. If desired, click on one or more of these links to see what's available to you



### 1.3.3 Administering BigInsights

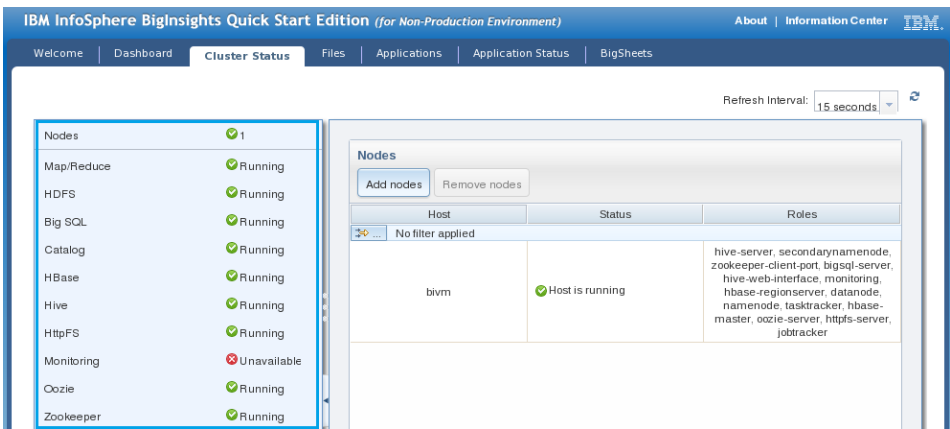
The Web console allows administrators to inspect the overall health of the system as well as perform basic functions, such as starting and stopping specific servers (or components), adding nodes to the cluster, and so on. You'll explore a subset of these capabilities here.

### 1.3.4 Inspecting the status of your cluster

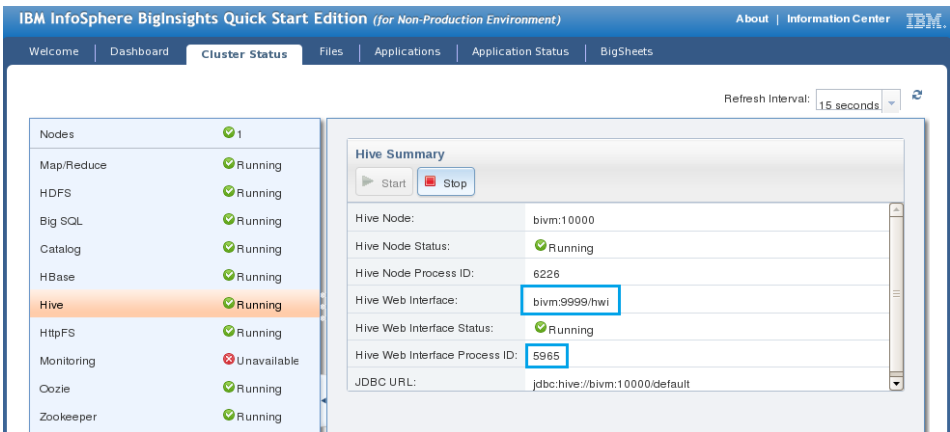
\_\_8. Click on the **Cluster Status** tab at the top of the page to return to the Cluster Status window



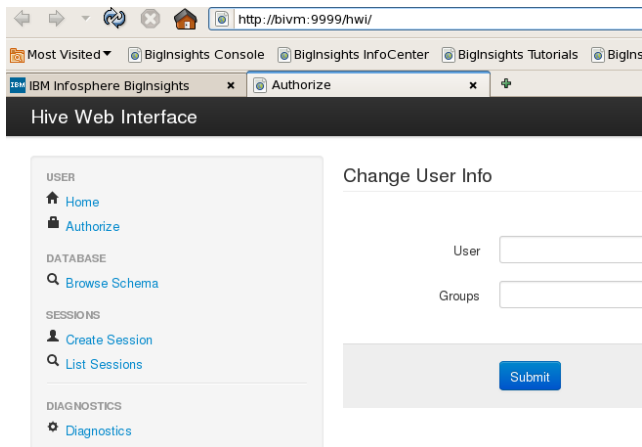
\_\_9. Inspect the overall status of your cluster. The figure below was taken on a single-node cluster that had several services running. One service – Monitoring -- was unavailable. (If you installed and started all BigInsights services on your cluster, your display will show all services to be running)



\_\_10. Click on the Hive service and note the detailed information provided for this service in the pane at right. From here, you can start or stop the hive service (or any service you select) depending on your needs. For example, you can see the URL for Hive's Web interface and its process ID.



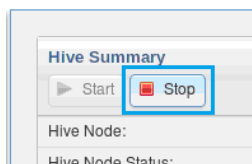
- \_\_\_11. Optionally, cut-and-paste the URL for Hive's Web interface into a new tab of your browser. You'll see an open source tool provided with Hive for administration purposes, as shown below.



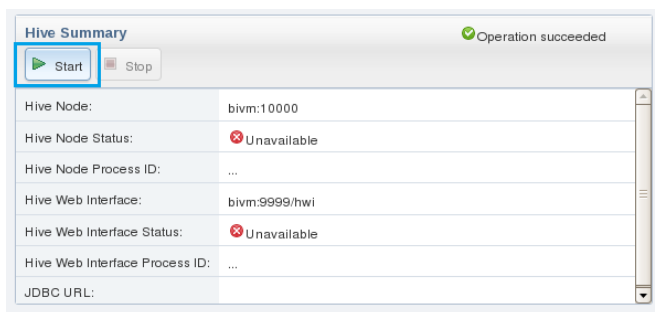
- \_\_\_12. Close this tab and return to the **Cluster Status** section of the BigInsights Web console

### 1.3.5 Starting and stopping a component

- \_\_\_13. If necessary, click on the Hive service to display its status.
- \_\_\_14. In the pane to the right (which displays the Hive status), click the red Stop button to stop the service



- \_\_\_15. When prompted to confirm that you want to stop the Hive service, click **OK** and wait for the operation to complete. The right pane should appear similar to the following image

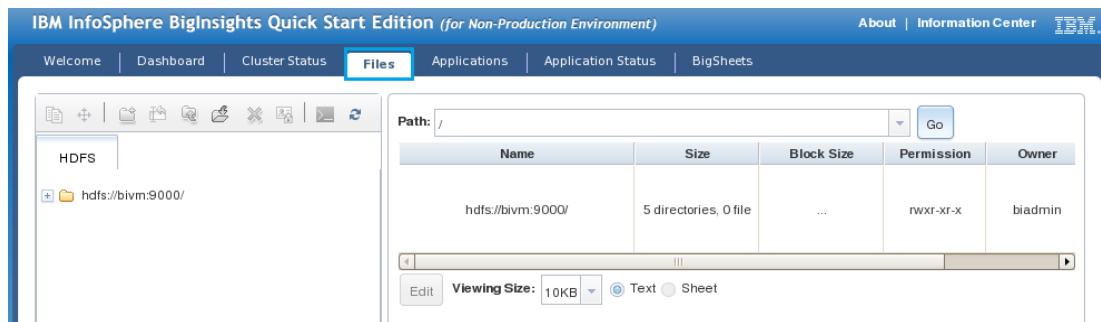


- \_\_\_16. Restart the Hive service by clicking on the green arrow just beneath the Hive Status heading. (See the previous figure.) When the operation completes, the Web console will indicate that Hive is running again, likely under a process ID that differs from the earlier Hive process ID shown at the beginning of this lab module. (You may need to use the Refresh button of your Web browser to reload information displayed in the left pane.)

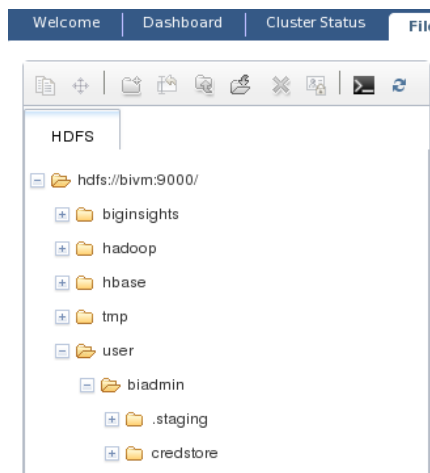
### 1.3.6 Working with Files

The **Files** tab of the console enables you to explore the contents of your file system, create new subdirectories, upload small files for test purposes, and perform other file-related functions. In this module, you'll learn how to perform such tasks against the Hadoop Distributed File System (HDFS) of BigInsights.

- \_\_\_17. Click on the **Files** tab of the console to begin exploring your distributed file system.



- \_\_\_18. Expand the directory tree shown in the pane at left (**/user/biadmin**). If you already uploaded files to HDFS, you'll be able to navigate through the directory to locate them.



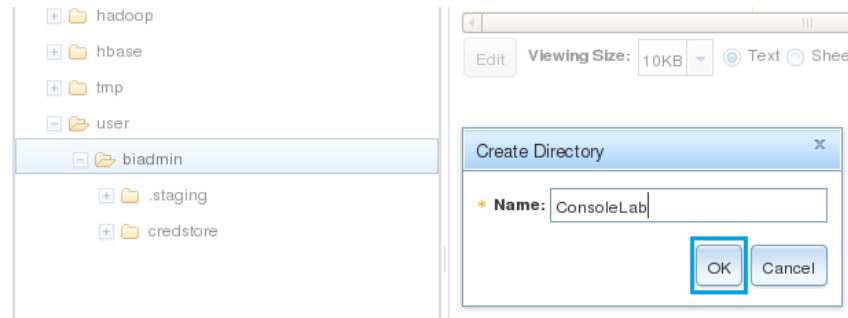
- \_\_\_19. Become familiar with the functions provided through the icons at the top of this pane, as we'll refer to some of these in subsequent sections of this module. Simply point your cursor at the icon to learn its function. From left to right, the icons enable you to Copy a file or directory, move a file, create a directory, rename, upload a file to HDFS, download a file from HDFS to your local file system, delete a file from HDFS, set permissions, open a command window to launch HDFS shell commands, and refresh the Web console page



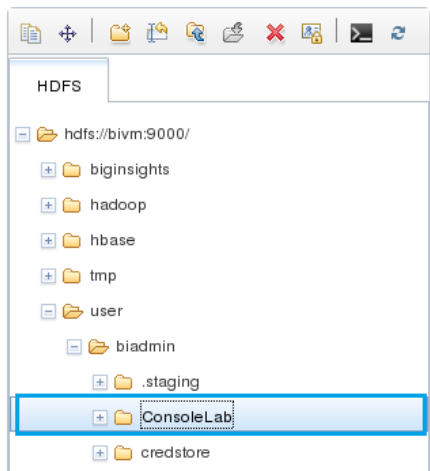
- \_\_20. Position your cursor on the **user/biadmin** directory and click the **Create Directory** icon to create a subdirectory for test purposes



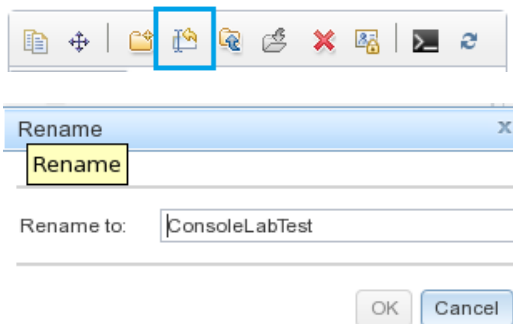
- \_\_21. When a pop-up window appears prompting you for a directory name, enter **ConsoleLab** and click **OK**



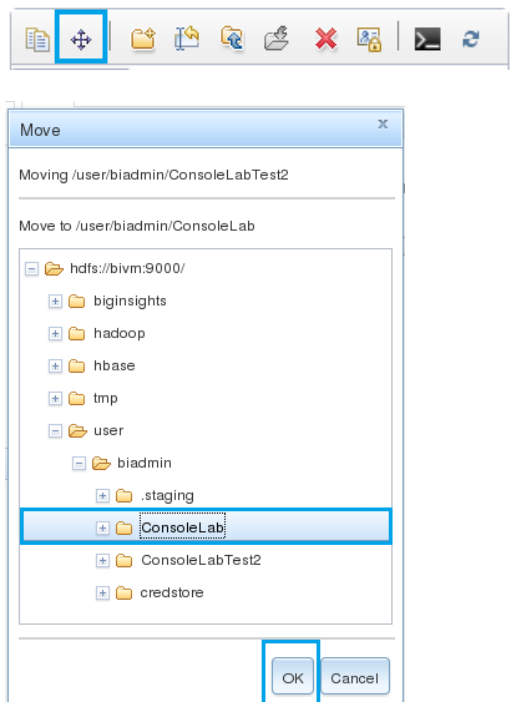
- \_\_22. Expand the directory hierarchy to verify that your new subdirectory was created.



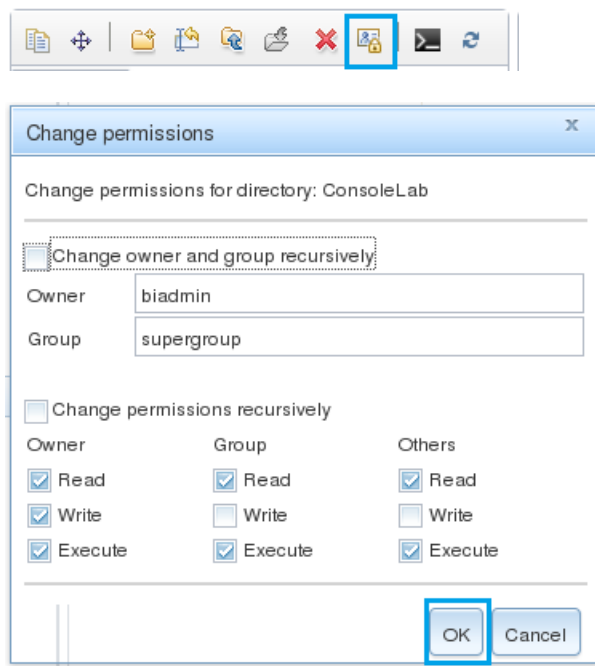
- \_\_23. Create another directory named ConsoleLabTest.
- \_\_24. Use the **Rename** icon to rename this directory to ConsoleLabTest2



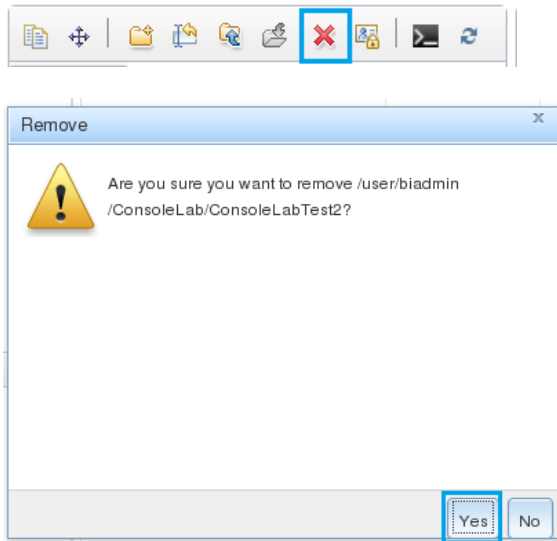
- \_\_\_25. Click the **Move** icon, when the pop up **Move** screen appears select the ConsoleLab directory and click **OK**.



- \_\_\_26. Using the **set permission** icon, you can change the permission settings for your directory. When finished click **OK**.



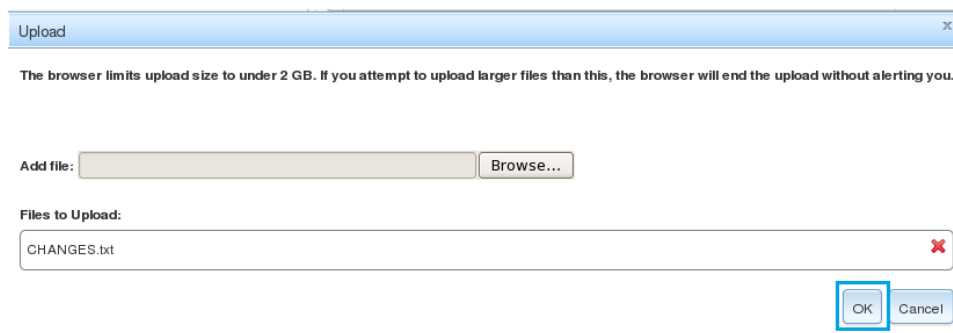
- \_\_27. While highlighting the ConsoleLabTest2 folder, select the **Remove** icon and delete the directory.



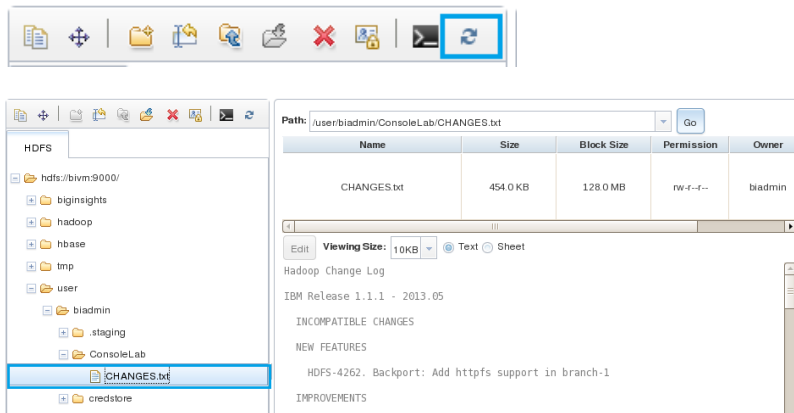
- \_\_28. Remain in the ConsoleLab directory, and click the **Upload** icon to upload a small sample file for test purposes.



- \_\_29. When the pop-up window appears, click the **Browse** button to browse your local file system for a sample file.
- \_\_30. Navigate through your local file system to the directory where BigInsights was installed. For the IBM-provided VMWare image, BigInsights is installed in file system: **/opt/ibm/biginsights**. Locate the **.../IHC** subdirectory and select the **CHANGES.txt** file. Click **Open**.
- \_\_31. Verify that the window displays the name of this file. Note that you can continue to **Browse** for additional files to upload and that you can delete files as upload targets from the displayed list. However, for this exercise, simply click **OK**



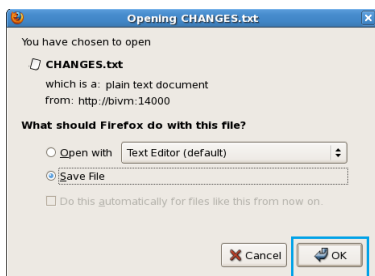
- \_\_32. When the upload completes, verify that the CHANGES.txt file appears in the directory tree at left. If it is not immediately visible click the refresh button. On the right, you should see a subset of the file's contents displayed in text format



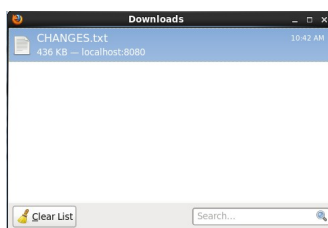
- \_\_33. Highlight the CHANGES.txt file in your ConsoleLab directory and click the **Download** button.



- \_\_34. When prompted, click the **Save File** button. Then select **OK**.



- \_\_35. If Firefox is set as default browser, the file will be saved to your user Downloads directory. For this exercise, the default directory location is fine



## 1.4 Summary

Congratulations! You're now familiar with the Hadoop Distributed File System. You know now how to manipulate files within by using the terminal and the BigInsights Web Console. You may move on to the next Unit.



## NOTES

[illegible]

## NOTES

[illegible]



© Copyright IBM Corporation 2013.

The information contained in these materials is provided for informational purposes only, and is provided AS IS without warranty of any kind, express or implied. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, these materials. Nothing contained in these materials is intended to, nor shall have the effect of, creating any warranties or representations from IBM or its suppliers or licensors, or altering the terms and conditions of the applicable license agreement governing the use of IBM software. References in these materials to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates. This information is based on current IBM product plans and strategy, which are subject to change by IBM without notice. Product release dates and/or capabilities referenced in these materials may change at any time at IBM's sole discretion based on market opportunities or other factors, and are not intended to be a commitment to future product or feature availability in any way.

IBM, the IBM logo and [ibm.com](http://ibm.com) are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).



Please Recycle