

Using BigSheets for Spreadsheet-like Analytics

Loading data into BigSheets

Contents

| | |
|---|-------------------------------------|
| LOADING DATA INTO BIGSHEETS | 4 |
| 1.1 GETTING STARTED | 5 |
| 1.2 LOADING DATA INTO BIGSHEETS FROM THE <i>FILES</i> TAB | 9 |
| 1.3 LOADING DATA INTO BIGSHEETS FROM THE <i>BIGSHEETS</i> TAB | ERROR! BOOKMARK NOT DEFINED. |
| 1.4 LOADING DATA INTO BIGSHEETS FROM THE DATA GENERATED BY AN APPLICATION | ERROR! BOOKMARK NOT DEFINED. |
| SUMMARY | 18 |

Loading data into BigSheets

Before you can work with BigSheets, you need to load the data that you want to work with into BigSheets. BigSheets is part of the IBM BigInsights product so you will be working with this for the lab exercise.

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

- Import data into a workbook
 - From the *Files* tab in the BigInsights Console
 - From the *BigSheets* tab in the BigInsights Console
 - Directly from an application on the BigInsights Console

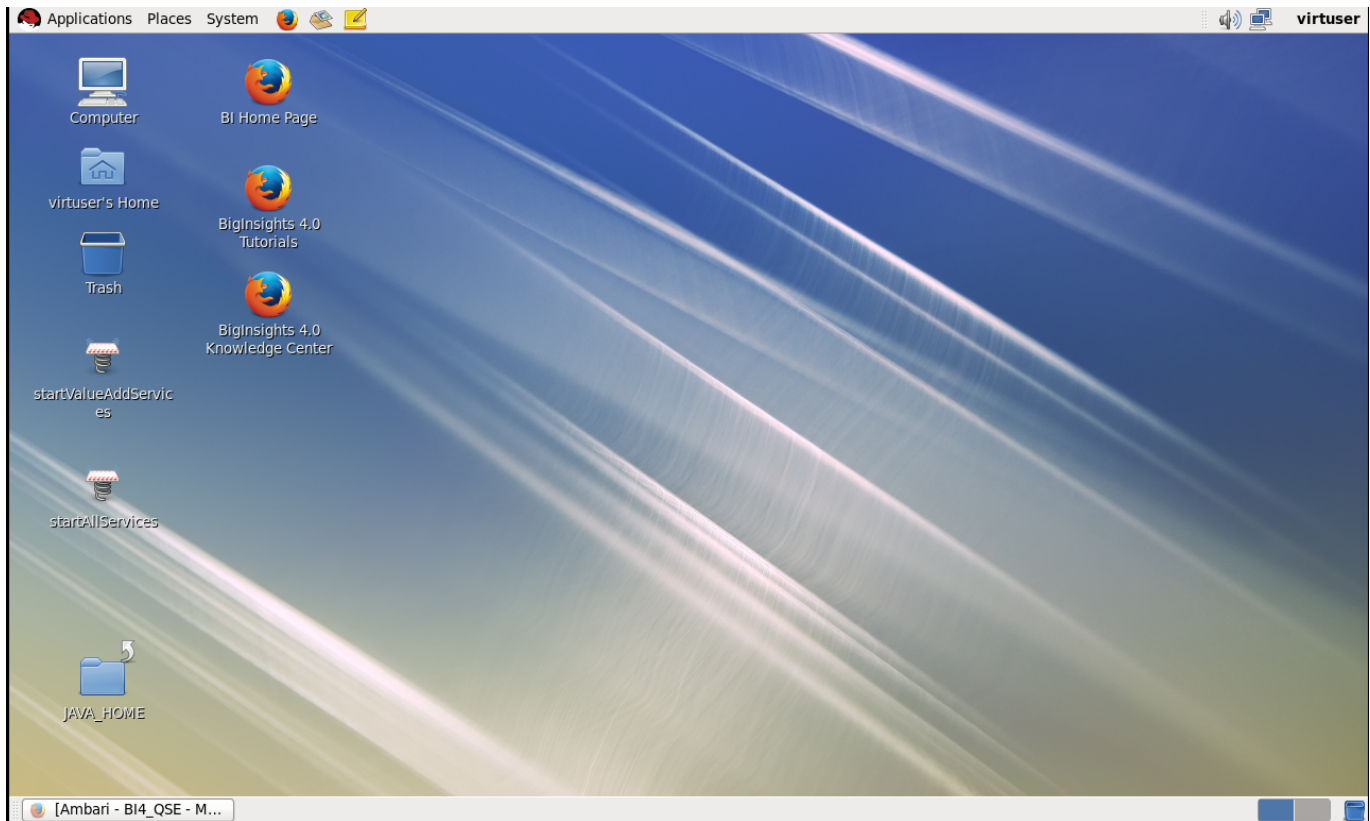
Allow 15-30 minutes to complete this section of lab.

Throughout this lab you will be using the following account login information:

| When to use: | Username | Password |
|---|----------|----------|
| Log in from the command-line to accept the licenses | root | password |
| Log in from the RHEL desktop | virtuser | password |
| Log in to the Ambari console | admin | admin |

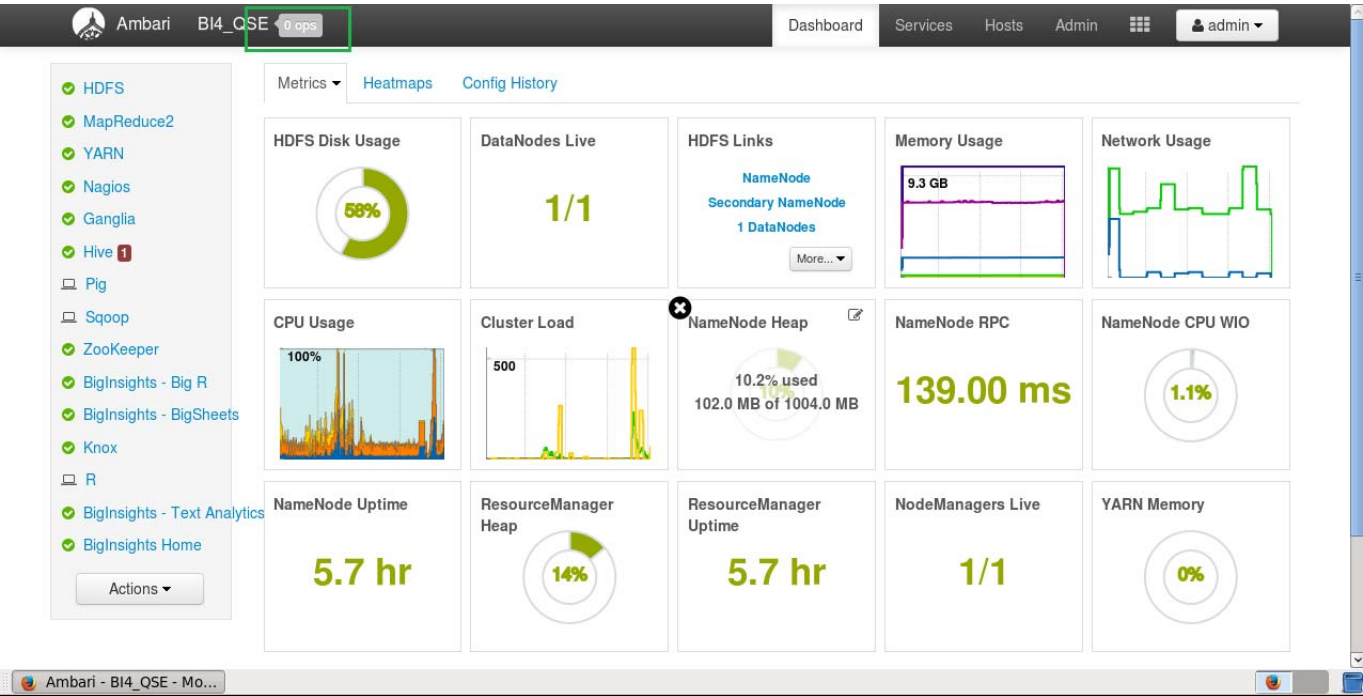
1.1 Getting Started

__1. After you log in, your screen should look similar to the one below.

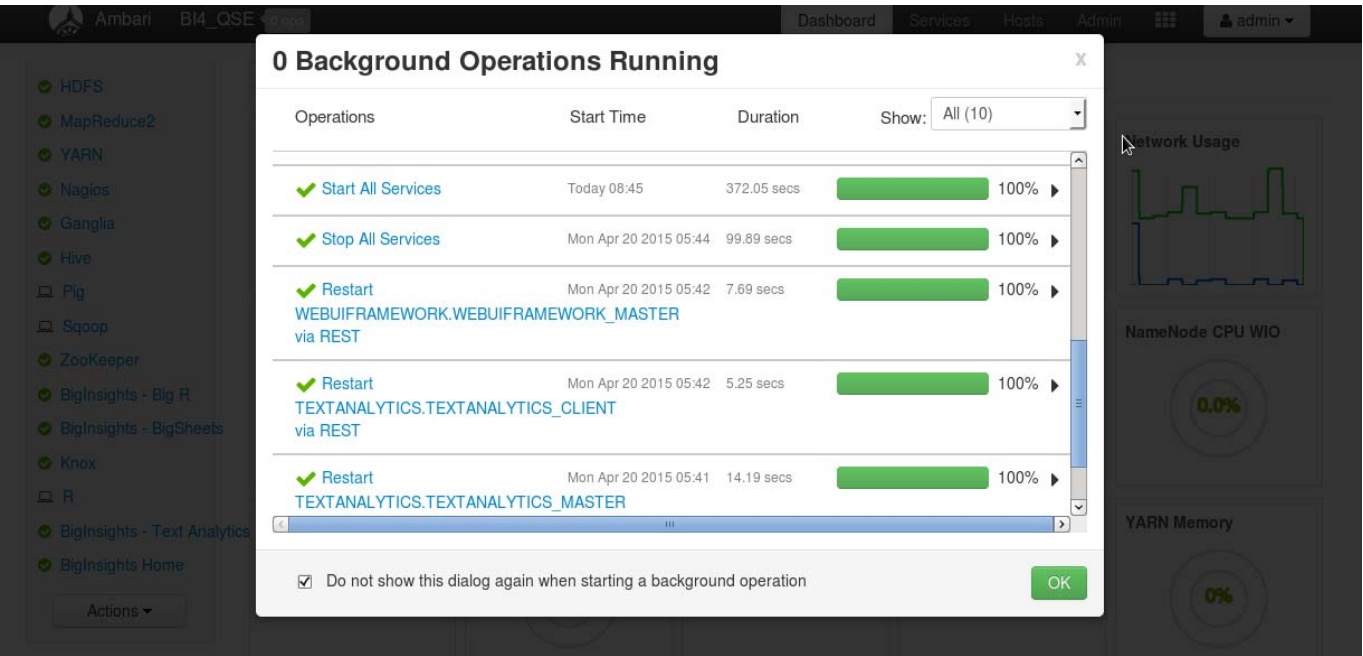


Before you can start working with BigSheets, you need to start up BigInsights. The Ambari console should have automatically started up for you. If not, you can start it up yourself. To check if the services have started, open up the Firefox browser and log in to the Ambari console. You will have to give it a few minutes for all the services to start up.

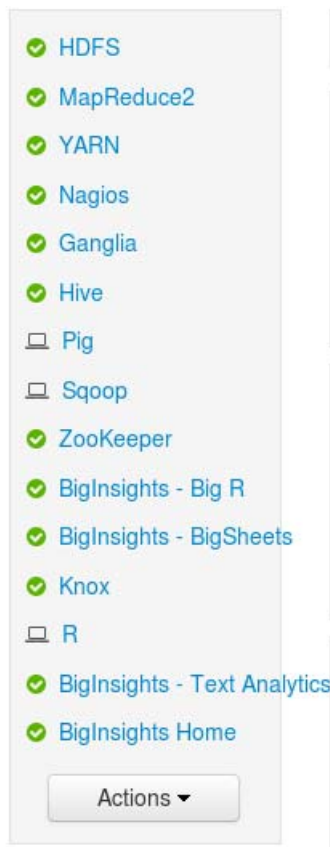
__2. Check to see the status of the startup by opening up the operations windows of the Ambari console:



__3. With the Operations dialog opened, you will see that the services are starting up:



__4. When all the services have started, the left panel should appear with all green check marks.



- ___5. Once that is up, you are ready to begin the lab. If not, use the appropriate buttons on the desktop to either "**startAllServices**" and the "**startValueAddServices**" (for BigSheets). You can also use the Services inside the Ambari console to start them up manually that way as well.
- ___6. Try these hints and tips if you have trouble with the installation of the VM Image.

1. Browser cannot load the Ambari user interface

If you have problems with the browser not being able to load the Ambari user interface, then the IP address of the machine might have changed due to a network connectivity reset. Network reset can happen during these conditions:

- Changing from wifi to wired or wired to wifi.
- Shutting down the system, and then re-establishing connection.
- Sending the machine into sleep mode. When it comes back up, the network IP address might be reset.

To resolve this issue, run the following scripts as the **root** user (root/password):

```
/home/virtuser/setHostname.sh
/home/virtuser/start-all.sh
```

Then, try to connect to `http://rvm.svl.ibm.com:8080`. You can also reboot the VM so that the IP address changes are updated as part of the initialization scripts.

2. Some of the BigInsights value-add services will not launch.
 1. From the Ambari web interface, click **Actions > Stop all services**.
 2. From the Ambari web interface, click **Actions > Start all services**.
 3. Run the **startValueAddServices** script that is available on the desktop of the Quick Start Edition.
 4. From the VM image desktop, click **BIHomePage** to launch the BigInsights Home web interface.

1.2 Loading data into BigSheets using HDFS commands

In this section of the exercise, you will be creating a workbook using CSV files. These files represent data from a fictitious company, The Great Outdoor Company. You should have loaded the labfiles directory on under /home/virtuser. If not, go back to that section on the BDU page for BigSheets to download the labfiles directory onto the VM.

- __1. Launch a terminal.
- __2. Type in the command to copy the labfiles directory into HDFS under /user/virtuser.

```
hdfs dfs -put /home/virtuser/labfiles /user/virtuser/
```

- __3. Let's take a moment to understand the files that we just uploaded to the HDFS.

```
hdfs dfs -ls /user/virtuser/labfiles
```

```
[virtuser@rvm Desktop]$ hdfs dfs -ls /user/virtuser/labfiles
Found 7 items
-rw-r--r--  3 virtuser hdfs      606 2015-05-04 11:19 /user/virtuser/labfiles
/RDBMS_data.csv
-rw-r--r--  3 virtuser hdfs  1416881 2015-05-04 11:19 /user/virtuser/labfiles
/blogs-data.txt
-rw-r--r--  3 virtuser hdfs    3389 2015-05-04 11:19 /user/virtuser/labfiles
/employee_state .csv
-rw-r--r--  3 virtuser hdfs  881781 2015-05-04 11:19 /user/virtuser/labfiles
/last_of_the_mohicans.txt
-rw-r--r--  3 virtuser hdfs  2634155 2015-05-04 11:19 /user/virtuser/labfiles
/news-data.txt
-rw-r--r--  3 virtuser hdfs    1386 2015-05-04 11:19 /user/virtuser/labfiles
/product.csv
-rw-r--r--  3 virtuser hdfs    1223 2015-05-04 11:19 /user/virtuser/labfiles
/sales.csv
[virtuser@rvm Desktop]$
```

RDBMS_data.csv → data from a RDBMS database

blogs-data.txt → data from the boardreader application

employee_state.csv → data for the employees' state

last_of_the_mohicans.txt → text file from the last of the mohicans

news-data.txt → some news blogs

product.csv → csv file of the products from the The Great Outdoor Company

sales.csv → csv file of the sales

1.3 Loading data into BigSheets using the BigInsights Home page

In this section of the exercise, you will be creating a workbook using CSV files. These files represent data from a fictitious company, The Great Outdoor Company.

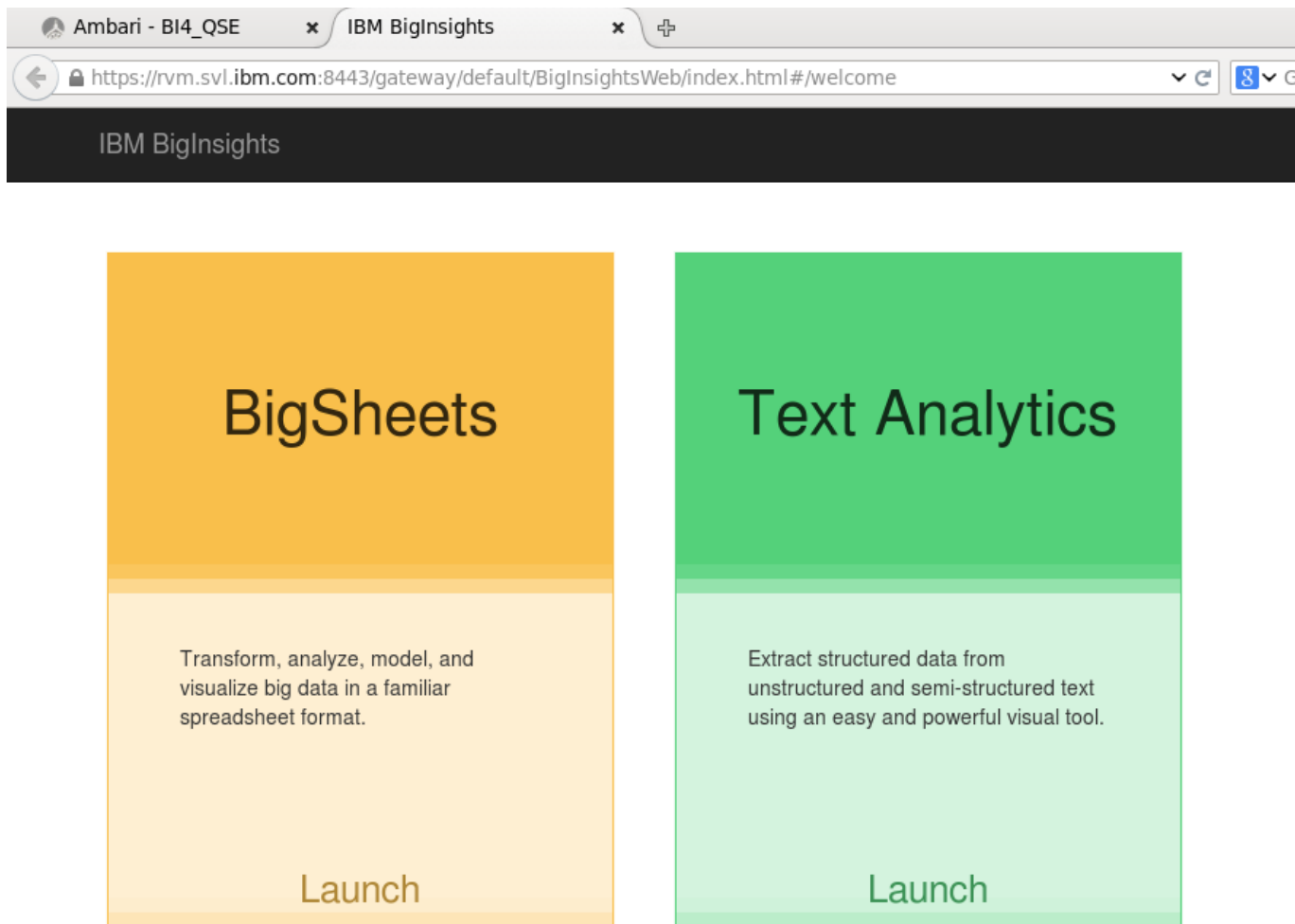
- ___4. Start the BigInsights Console by double-clicking on the *Web Console* icon on the desktop.



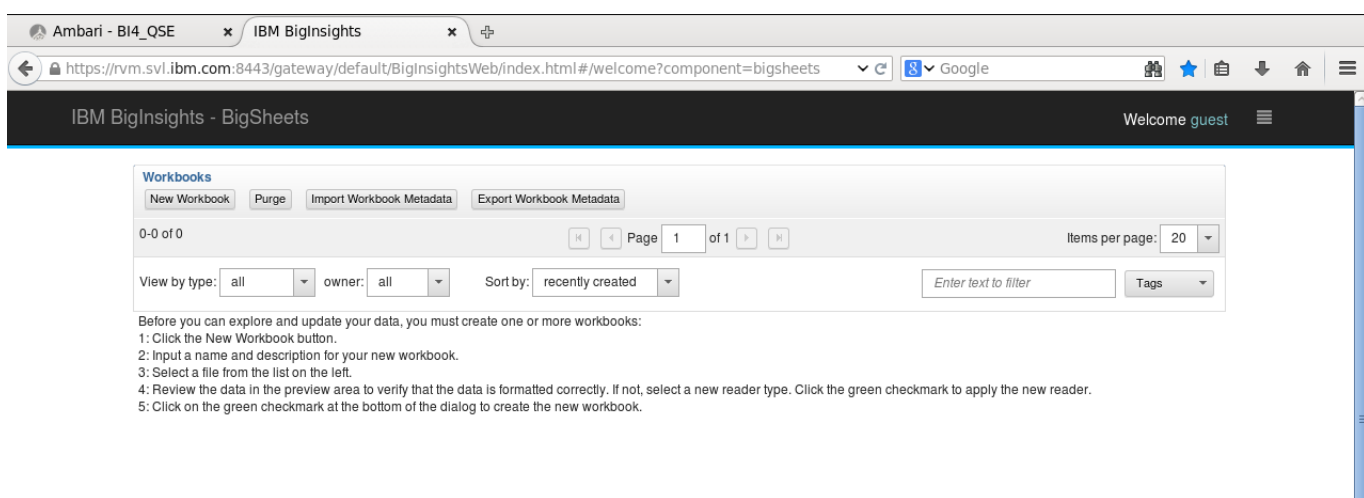
- ___5. If it doesn't start up, you need to start the **Demo LDAP** via the **Knox** service:

The screenshot displays the BigInsights console interface. On the left, a sidebar lists various services: HDFS, MapReduce2, YARN, Nagios, Ganglia, Hive, Pig, Sqoop, ZooKeeper, BigInsights - Big R, BigInsights - BigSheets, Knox (highlighted), R, BigInsights - Text Analytics, and BigInsights Home. An 'Actions' dropdown is at the bottom of the sidebar. The main panel shows the 'Summary' tab for the 'Knox Gateway' service, which is marked as 'Started'. To the right, an 'Alerts and Health Checks' section shows a green checkmark and the text 'Knox Gateway process TCP OK - 0.023 second response'. A 'Service Actions' dropdown menu is open, showing options: Start, Stop, Restart All, Run Service Check, Turn On Maintenance Mode, Start Demo LDAP (highlighted), and Stop Demo LDAP.

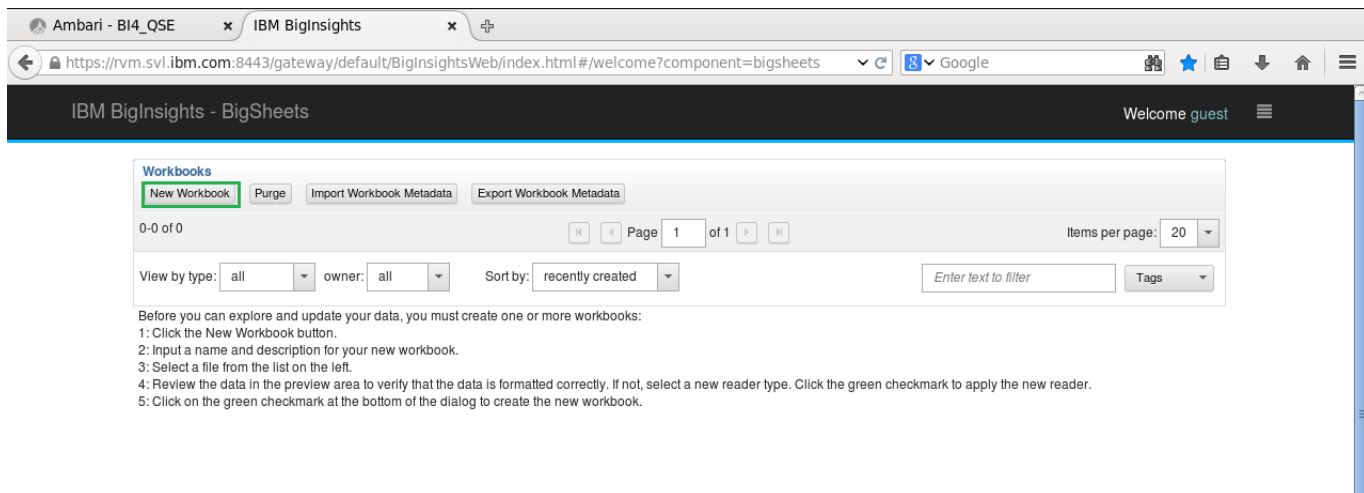
- ___6. The login to the BigInsights home should be saved in the browser. If not, use **guest/guest-password** as the login.



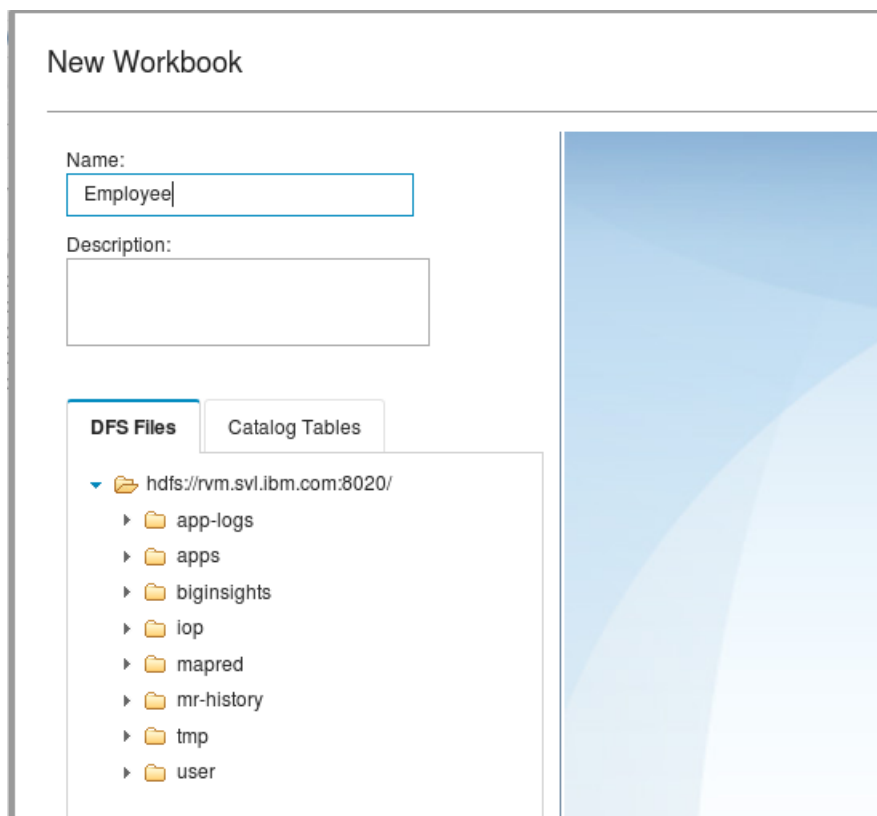
__7. Click the BigSheets to launch it.



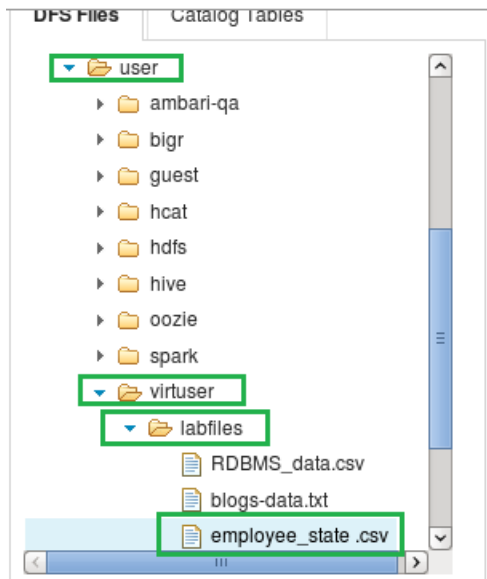
__8. Create a new workbook using the *employee_state.csv*. Click the New Workbook pushbutton.



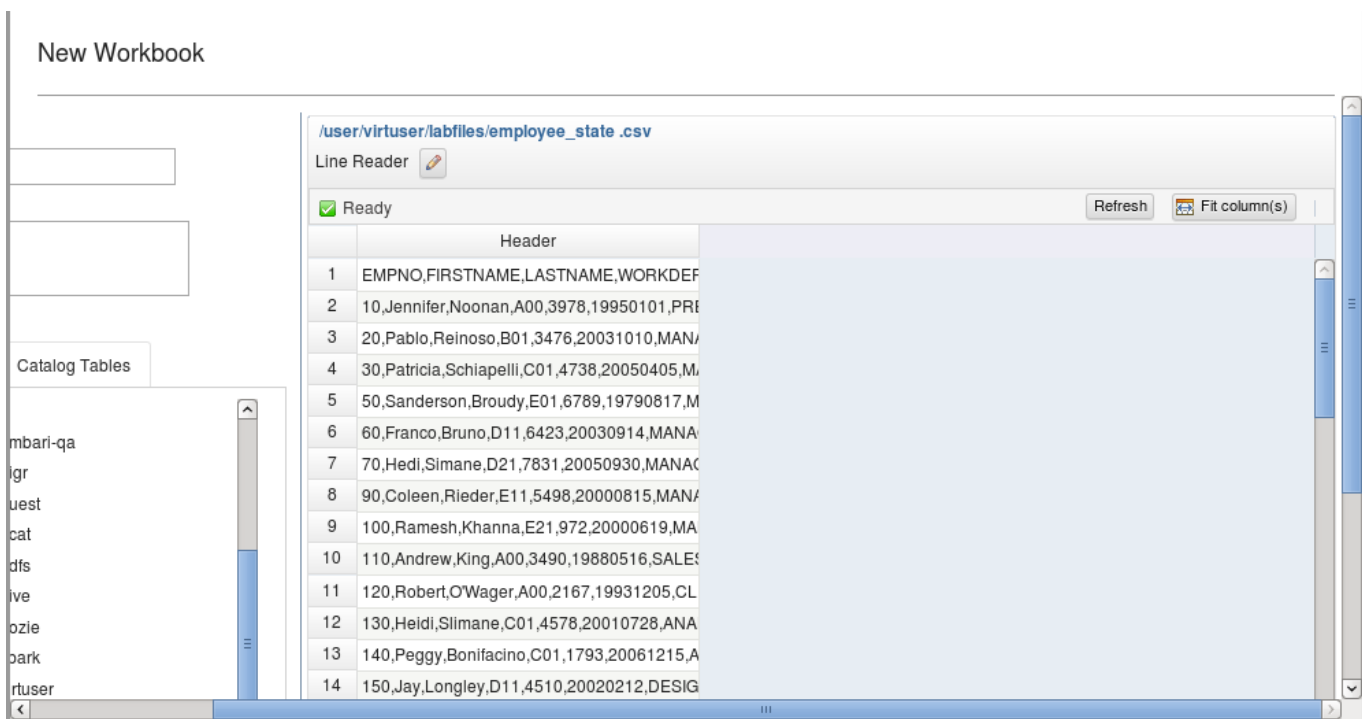
__9. In the dialog that pops up, give the workbook the name *Employee*.



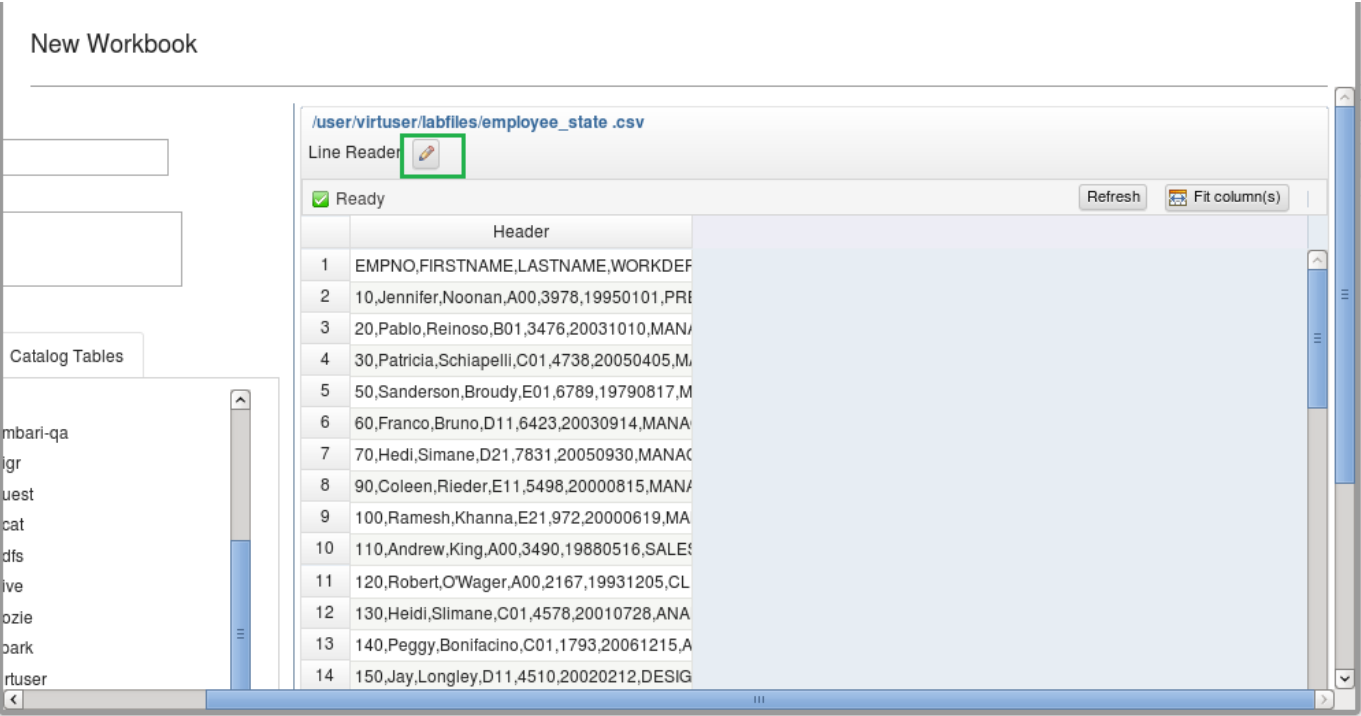
__10. Scroll down further, under the DFS Files tab, and select the *employee_state.csv* file under */user/virtuser/labfiles/employee_state.csv*



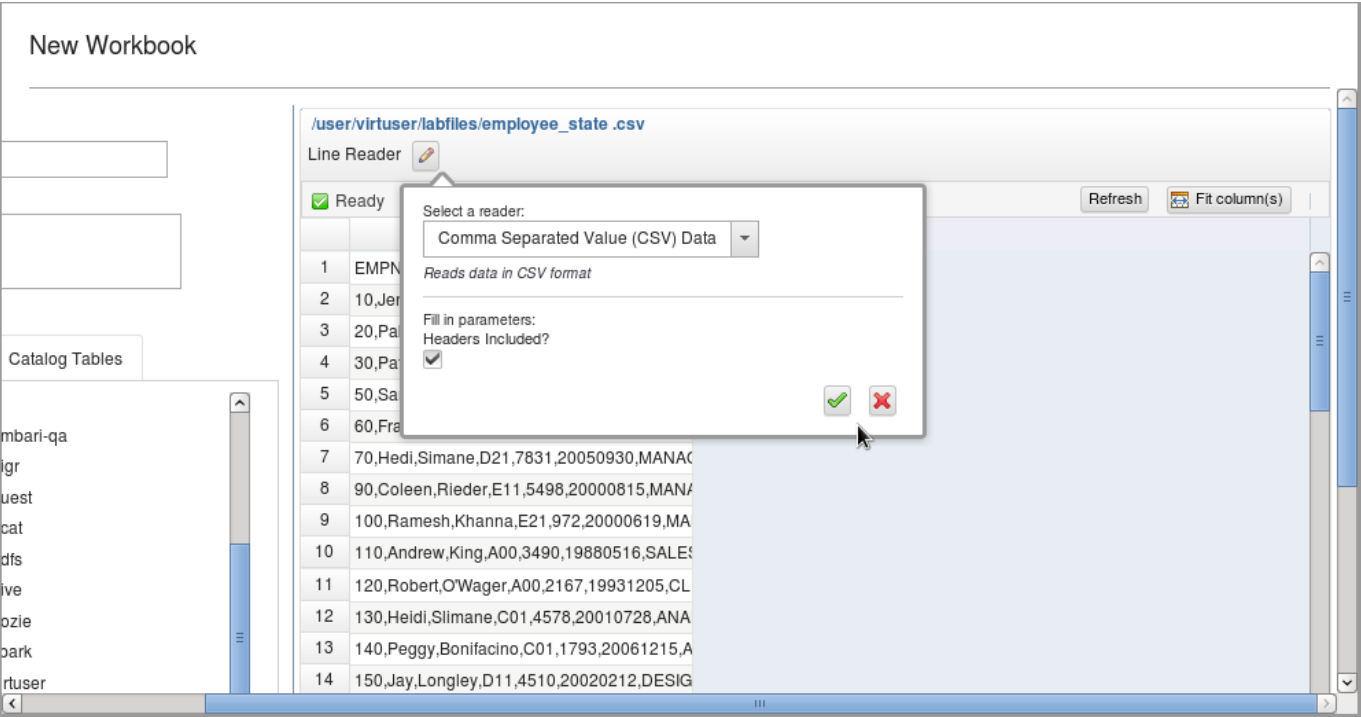
- ___11. The contents of *employee_state.csv* are displayed in the right frame in a text format. Scroll up and right to center the preview pane:



- ___12. The data has been read by the BigSheets' line reader. Unfortunately, the data is in a comma-separated format, so you need to specify a different BigSheets reader. Click the **pencil** icon.



- ___13. From the *Select a reader* drop-down box, select *Comma Separated Value (CSV) Data*. Since the data includes header information, make sure the *Headers included?* checkbox is checked. Click the **green check mark**.



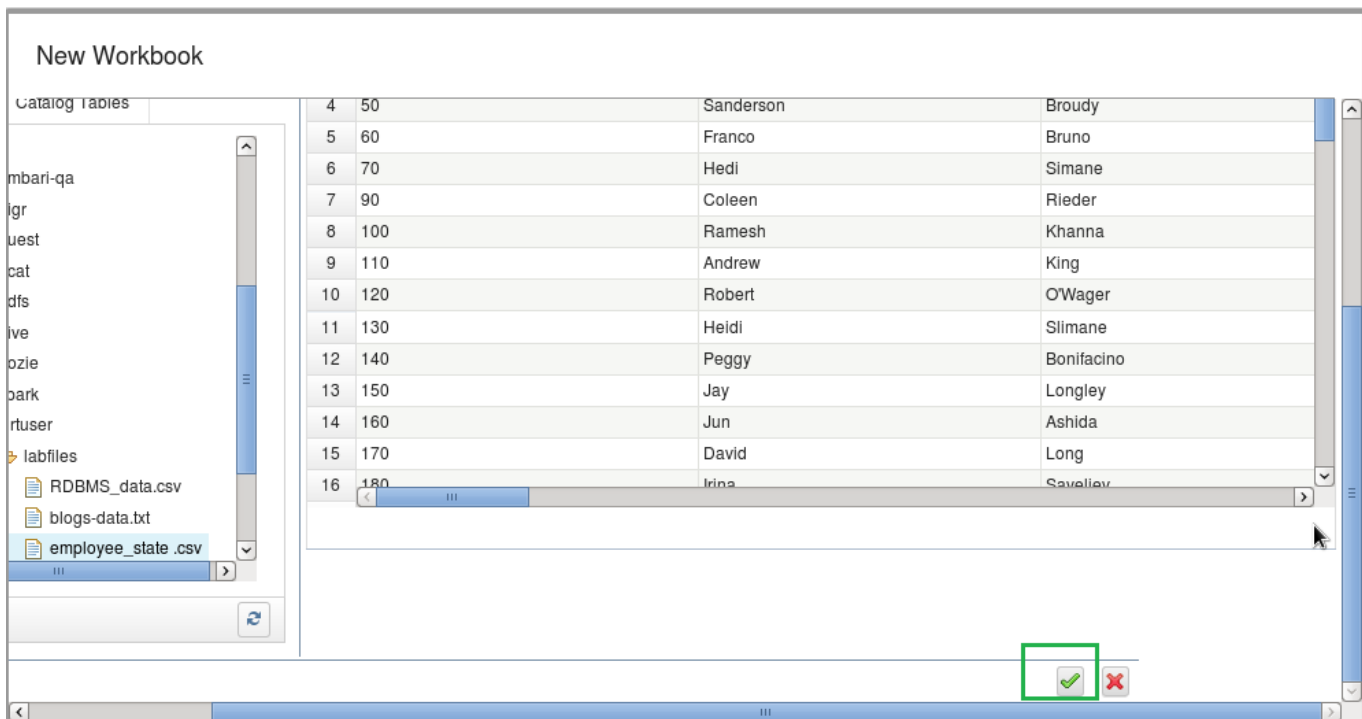
- ___14. The data is now formatted correctly.

/user/virtuser/labfiles/employee_state.csv
Comma Separated Value (CSV) Data

☒ Ready Refresh Fit column(s)

| | EMPNO | FIRSTNAME | LASTNAME |
|----|-------|-----------|------------|
| 1 | 10 | Jennifer | Noonan |
| 2 | 20 | Pablo | Reinoso |
| 3 | 30 | Patricia | Schiapelli |
| 4 | 50 | Sanderson | Broudy |
| 5 | 60 | Franco | Bruno |
| 6 | 70 | Hedi | Simane |
| 7 | 90 | Coleen | Rieder |
| 8 | 100 | Ramesh | Khanna |
| 9 | 110 | Andrew | King |
| 10 | 120 | Robert | O'Wager |
| 11 | 130 | Heidi | Slimane |
| 12 | 140 | Peggy | Bonifacino |
| 13 | 150 | Jay | Longley |
| 14 | 160 | Jun | Ashida |

- ___15. Scroll to the bottom of the New Workbook dialog to Save the workbook. Click the green checkmark.



This time, use the same process for loading in the products csv file. We're not going to do anything with the Employee workbook just yet.

- ___16. Click the Workbooks breadcrumb link near the top to get back to the BigSheets home.

IBM BigInsights - BigSheets

Welcome guest

Workbooks View Results

Employee

Delete Add chart Employee : Build new workbook

Ready Refresh Fit column(s) Create Table Export data Export metadata Run Stop 100%

| | EMPNO | FIRSTNAME | LASTNAME | WORKDEPT | PHONENO |
|----|-------|------------|------------|----------|---------|
| 1 | 10 | Jennifer | Noonan | A00 | 3978 |
| 2 | 20 | Pablo | Reinoso | B01 | 3476 |
| 3 | 30 | Patricia | Schiapelli | C01 | 4738 |
| 4 | 50 | Sanderson | Broudy | E01 | 6789 |
| 5 | 60 | Franco | Bruno | D11 | 6423 |
| 6 | 70 | Hedl | Simane | D21 | 7831 |
| 7 | 90 | Coleen | Rieder | E11 | 5498 |
| 8 | 100 | Ramesh | Khanna | E21 | 972 |
| 9 | 110 | Andrew | King | A00 | 3490 |
| 10 | 120 | Robert | O'Wager | A00 | 2167 |
| 11 | 130 | Heidi | Simane | C01 | 4578 |
| 12 | 140 | Peggy | Bonifacino | C01 | 1793 |
| 13 | 150 | Jay | Longley | D11 | 4510 |
| 14 | 160 | Jun | Ashida | D11 | 3782 |
| 15 | 170 | David | Long | D11 | 2890 |
| 16 | 180 | Irina | Saveliev | D11 | 1682 |
| 17 | 190 | Shih-Sheng | Yang | D11 | 2986 |

Preview of 42 rows from ??? Prev Next

- __17. Click the **New Workbook** pushbutton.
- __18. In the Name field, type in **Products**.
- __19. Drill down the *DFS File* to */user/virtuser/labfiles/*. Select the *product.csv* file and see that its content is displayed on the right.
- __20. Similar to the *employee_state.csv* file that you did earlier, the format is in CSV so you need to specify the appropriate line reader. Select the **Pencil** icon and choose the *Comma Separated Value (CSV) Data*. The file also includes header, so make sure the *Headers included?* checkbox is checked. Click the **green check mark**.
- __21. Scroll down to the bottom and click the **green check mark** to create the new *Products* workbook.
- __22. You are going to do the same steps for the *sales.csv* file.
- __23. Click the **New Workbook** pushbutton.
- __24. In the Name field, type in **Sales**.
- __25. Drill down the *DFS File* to */user/virtuser/labfiles/*. Select the *sales.csv* file and see that its content is displayed on the right.

- ___26. Select the **Pencil** icon and choose the *Comma Separated Value (CSV) Data*. The file also includes header, so make sure the *Headers included?* checkbox is checked. Click the **green check mark**.
- ___27. Scroll down to the bottom and click the **green check mark** to create the new *Sales* workbook.

Summary

Having completed this exercise, you should now be able to load data into BigSheets using the Hadoop commands along with the BigInsights Home page to create worksheets. You specified the CSV reader to load in CSV files.

NOTES

NOTES



© 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 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.



Please Recycle
