

## Demonstration 1

Import data into a workbook.

At the end of this demonstration, you should be able to:

- Load data into various master workbooks

### *Demonstration 1: Importing data into a workbook*

## Demonstration 1: Importing data into a workbook

### Purpose:

You will load data into various BigSheets master workbooks.

User/Password: **biadmin/biadmin**  
**Root/dalvm3**

Service Password: **ibm2blue**

### Task 1. Creating a workbook from a web crawler application.

In this task, you will be creating a workbook using the results of a Web Crawler application. The crawler was directed to extract information from a website that dealt with patents. Essentially the web crawler looked at a site that had a list of names. Each name is a hyperlink to a page that lists the patents for that person.

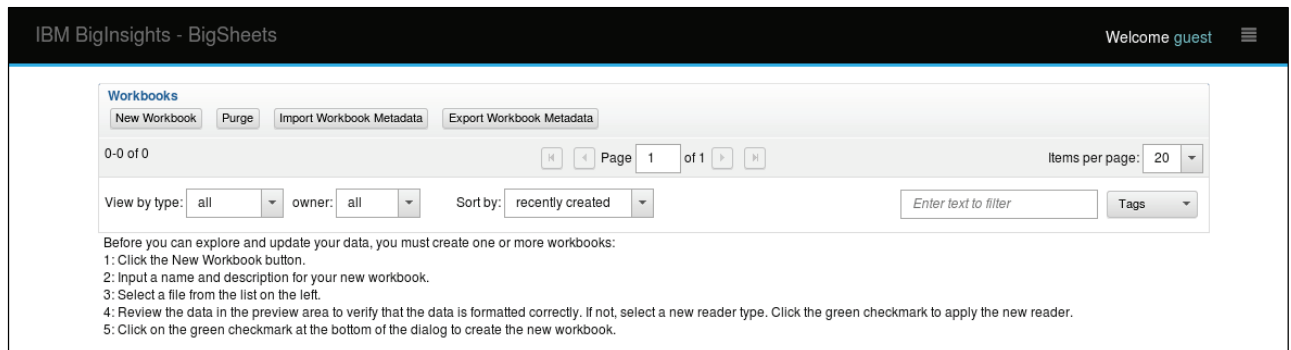
1. Launch **Firefox**, or open up a new tab in the existing browser and go to the following website:  
<http://www.ibm.com/software/ebusiness/jstart/bigsheets/demo/Patents.html>  
There you see the list of names.
2. Click on any name and it takes you to a page that lists all of the patents registered to that individual. This is to give you a frame of reference when doing this exercise.
3. You can close that tab.
4. Open a new terminal or use an existing one. To open, right-click anywhere on the desktop and click **Open in Terminal**.
5. Upload the results of the web crawler run into the HDFS.




```
hdfs dfs -put
/home/biadmin/labfiles/bigsheets/PatentCrawler_data.csv
/user/biadmin/
```

This material is meant for IBM Academic Initiative use only. NOT FOR RESALE

- Go to the **BigInsights Home** page in the Firefox browser and click on the **BigSheets** link (last step from the previous demonstration) to load the BigSheets landing page.

You should be here now:



- Click the **New Workbook** to bring up the dialog.
- For the **Name** field, type **PatentCrawler**.
- For the **Description** field, type **Results from Web Crawler**.
- On the **DFS Files** tab, drill down to **/user/biadmin/** and then select **PatentCrawler\_data.csv**.  
On the Preview pane, you will see that the default Line Reader was used. The file format is CSV, so you will select the CSV reader.
- Click **Edit Workbook reader**  and then select the **Comma Separated Value (CSV) data**.
- Uncheck** the "Headers Included?" checkbox and then click **Apply settings**  to change the reader type.
- Click **Save workbook**  to create the *PatentCrawler* workbook.
- You will work with this workbook in a later lab exercise. For now, you are done with this task.

This material is meant for IBM Academic Initiative use only. NOT FOR RESALE

## Task 2. Creating a workbook from the results of a WordCount application.

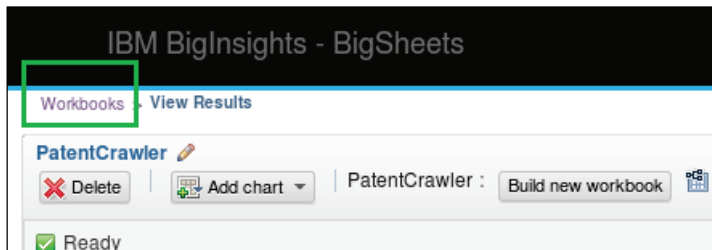
In this task, you will be creating a workbook from the results of a WordCount application that was ran against the file, `last_of_the_mohicans.txt`, which is located on your local system under `/home/biadmin/labfiles/bigsheets`. The output file from the WordCount run is also located under the same directory -- `part-r-00000`. You will load this file to the HDFS in order to create a workbook from it.

1. Use an existing terminal, or open a new one.


2. Upload `part-r-00000` to the HDFS.

```
hdfs dfs -put /home/biadmin/labfiles/bigsheets/
part-r-00000 /user/biadmin
```

3. Go back to the **BigSheets Home** page by clicking on the **Workbooks** breadcrumb.



4. Click the **New Workbook** to bring up the dialog.
5. In the *Name* field, type **Wordcount**.
6. In the *Description* field, type **Results from WordCount application**.
7. On the **DFS Files** tab, drill down to `/user/biadmin/` and select **part-r-00000**. On the Preview pane, you will see that the default Line Reader was used. This is sufficient for this file.

8. Click **Save workbook**  to create the Wordcount workbook. You will work with this workbook in a later lab exercise. For now, you are done with this task.

## Task 3. Creating additional master workbooks.

In this task, you will create workbooks based off of the files you had uploaded to the HDFS in the first lab exercise. You will create three notebooks from the CSV files.

1. Go to the **BigSheets Home** page, and then click the **Workbooks** breadcrumb.
2. Click **New Workbook**.



The New Workbook window appears.

This material is meant for IBM Academic Initiative use only. NOT FOR RESALE

3. In the **Name** field, type **Employees**.
4. In the **Description** field, type **Employees state data**.
5. On the **DFS Files** tab, navigate to your user directory (/user/biadmin/) and select the **employee\_state.csv** file.

The contents of the employee\_state.csv are displayed in the right frame in a text format.

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.

6. Click **Edit workbook reader** .
7. Select **Comma-Separated-Value (CSV) Data**, ensure the **Headers included?** checkbox is selected and then click **Set reader** .
8. Create workbooks for the **product.csv** and **sales.csv** files.
9. Click **Save workbook**.

Once you have the three workbooks created, you are done with this task. You will get to work with these workbooks in a later exercise.

**Results:**

**You loaded data into various BigSheets master workbooks.**