ADSA Hadoop Introduction Workshop
10/2/2015

Sample Program 1: Pig Word Count

1. With the HortonWorks 2.3 environment running, open a browser to Ambari at:
   http://127.0.0.1:8080/#/login
2. Login to Ambari with username **admin** and password **admin**
3. Ambari will show if all the services are running or not. If a service is not running and shows a red alert next to its name, click on each service name that has a red alert and correct the issue. Usually, the service needs to be started. Use the Service Action drop down menu (located in the upper right corner of the page) to start the service.
4. If you start a service, please wait until the service completely starts. A progress bar will be displayed.


The red alerts should disappear on their own after a short time. You should wait until all the alerts are gone. (Typically, you need to wait less than a minute for the alerts to disappear.) The services HDFS, MapReduce2, YARN, Hive, Oozie, and Zookeeper need to have green checkmarks shown next to their name (on the left hand side of the page). This will indicate all the services running properly

Here is a screen snapshot of Ambari with no alerts.
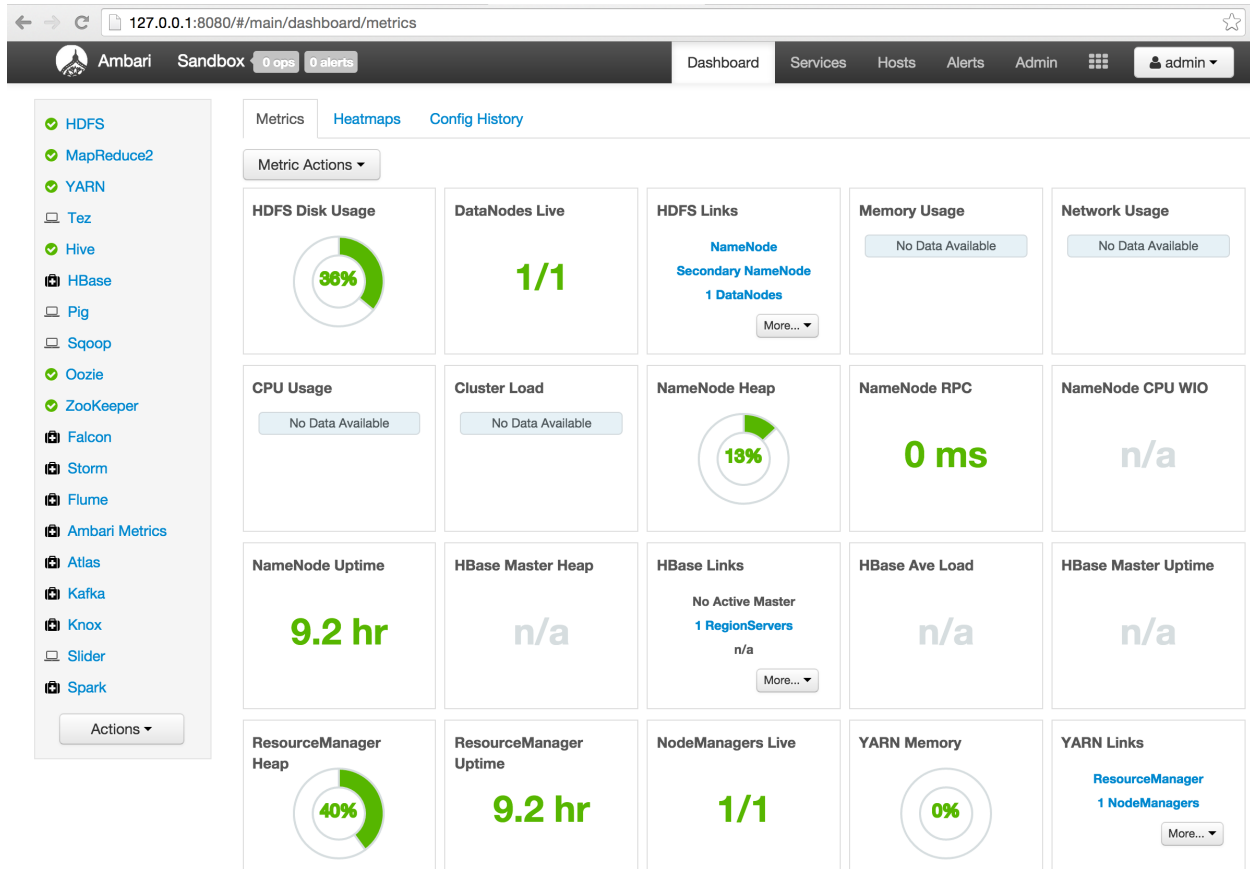
5. With all the services properly running, go to the web interface at http://127.0.0.1:8000/about/
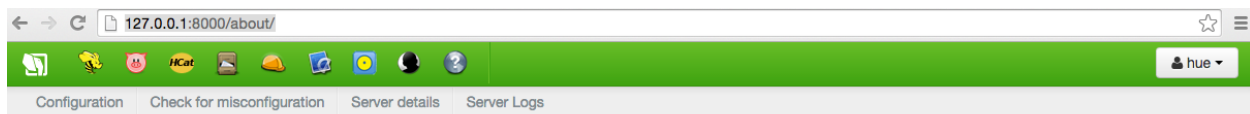
Hue

Component	Version	
Hue	2.6.1-2557	
HDP	2.3.0	
Hadoop	2.7.1	
Pig	0.15.0	
Hive-Hcatalog	1.2.1	
Oozie	4.2.0	
Ambari	2.1-1470	Disable
HBase	1.1.1	
Knox	0.6.0	
Storm	0.10.0	
Falcon	0.6.1	
Sandbox Build	f1dc3df 09:23 03-04-15	

6. Add a some files to the environment.
    - In the menu bar, click on the File Browser (File cabinet icon).
    - Create a new folder (named test)
    - Use the Upload button to import the data files
        - book.txt
        - mary.txt

7. Run the pig count words example.
    - In the menu bar, click on the Pig (Pig face icon)
    - Cut and paste the text from the **countwords.pig.txt** file into the script window.
    - Edit the path to the file on line 8 so it matches your location
        - '/user/hue/test/mary.txt'
    - Give the script a title (countwords.pig) and click on Save to save the file
    - Run the script by clicking on the Execute button
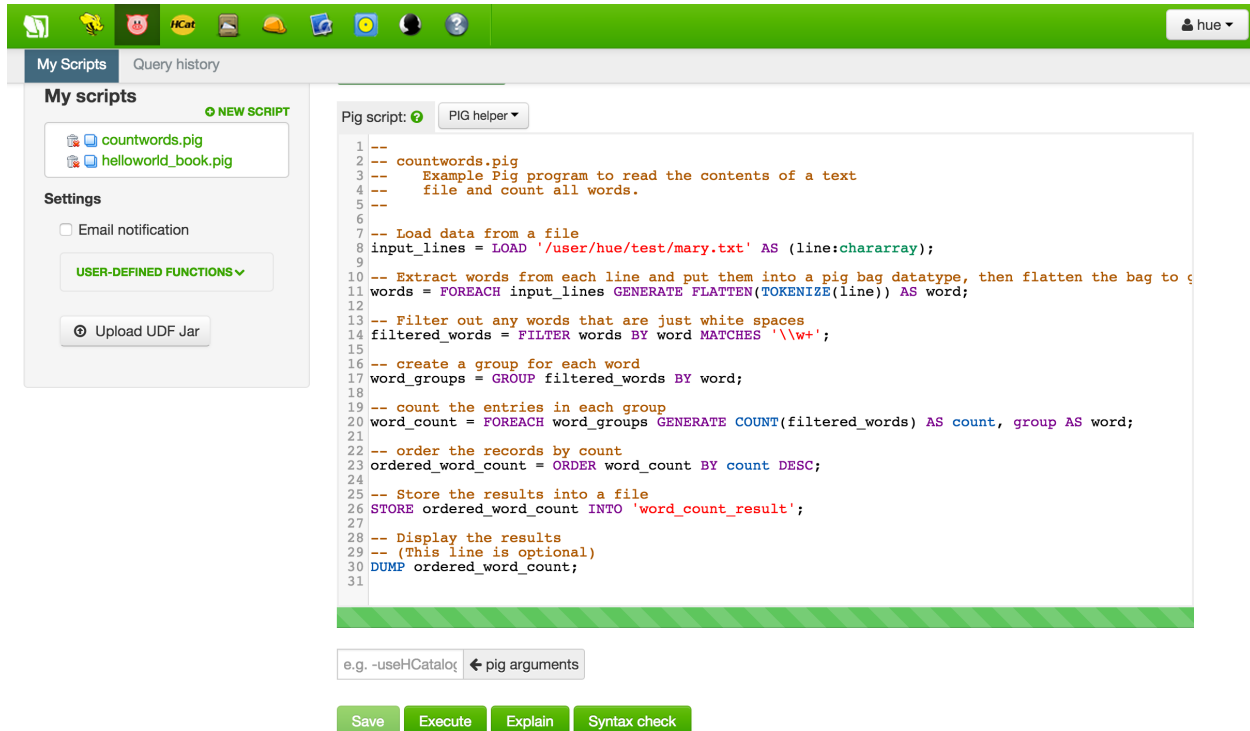
    -
    - You will see a progress bar at the bottom of the screen and a message that the job has started.
    - The job may take several minutes to complete. When the job completes, the Kill Job button will be replaced with the Execute Explain Syntax check buttons.
    - After the job has completed, Scroll down to view the output.

Here are some of the results:

Save   Execute   Explain   Syntax check

```
(8,the)
(5,lamb)
(5,mary)
(4,and)
(3,was)
(3,it)
(3,to)
(2,a)
(2,children)
(2,teacher)
(2,school)
(2,why)
(2,did)
(2,so)
(1,which)
(1,still)
(1,reply)
(1,loves)
(1,laugh)
(1,eager)
(1,about)
(1,went)
(1,till)
(1,that)
```
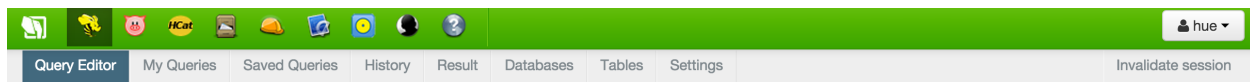
Sample Program 2: Pig Hello World

1. Start with the HortonWorks web interface with all the services properly running:
   http://127.0.0.1:8000/about/
2. Run the pig hello world example.
   - In the menu bar, click on the Pig (Pig face icon)
   - Cut and paste the text from the **helloworld.pig.txt** file into the script window.
   - Edit the path to the file on line 8 so it matches your location
     - '/user/hue/test/book.txt'
   - Give the script a title (**helloworld.pig**) and click on Save to save the file.
   - Run the script by clicking on the Execute button.
3. After the script completes, scroll to the last page of the output:
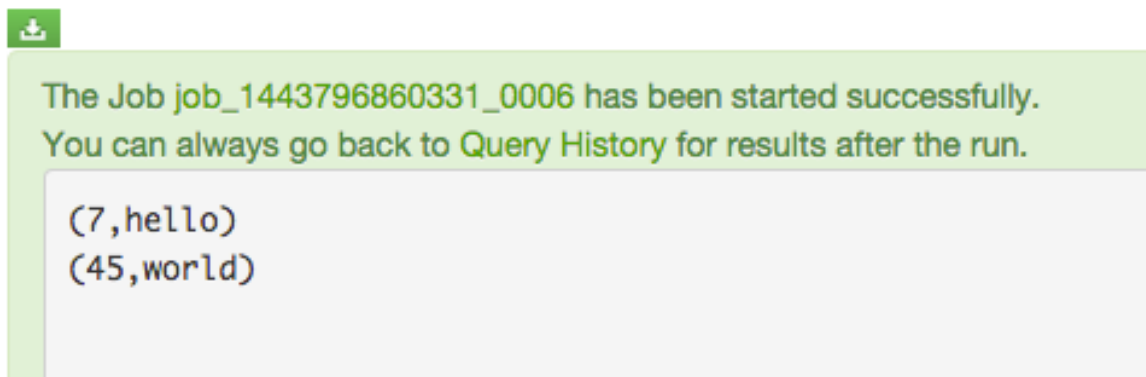
Sample Program 3: Hive Word Count

1. Go to the HortonWorks web interface, http://127.0.0.1:8000/about/
2. In the menu, click on the File Browser icon.
3. Upload the **book.txt** file in the /user/hue/test/ directory.
4. In the menu, click on the Hive icon (yellow bee-elephant).
5. Cut and paste the Hive program (**helloworld.hive.txt**) into the Query Editor.
6. Click on Execute to run the script. Wait for the script to complete.
7. Scroll through the results when they are displayed.

			hue ▾
Query Editor	My Queries	Saved Queries	History	Result	Databases	Tables	Settings	Invalidate session

Query Results: Hive hello world 3

Results   Query   Log   Columns

DOWNLOADS ‹

Download as CSV

Download as XLSX

Save

Did you know? If the result contains a large number of columns, click a row to select a column to jump to. As you type into the field, a drop-down list displays column names that match the string.   ✕

▲	_c0	word
85	3	Answer
86	1	Antonette
87	4	Any
88	3	Anybody
89	1	Anyhow
90	1	Anyway
91	2	Anyways
92	2	Apthorps
93	1	Apthorpswhich
94	1	Arabbut
95	1	Araminta
96	2	Archibald
97	13	Archive
98	2	Are
99	1	Aristocracy

Next Page →