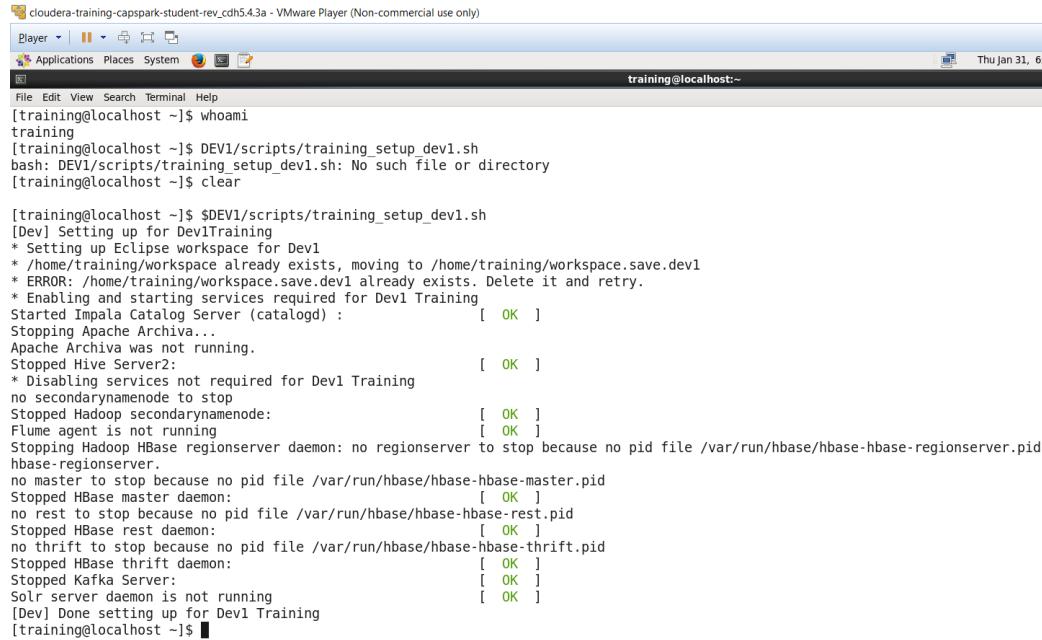


## Assignment 1 Big Data

### Access HDFS with Command line and Hue

- 1) The setup script is run before running other scripts



```
cloudera-training-capspark-student-rev_cdh5.4.3a - VMware Player (Non-commercial use only)
File Edit View Search Terminal Help
[training@localhost ~]$ whoami
training
[training@localhost ~]$ DEV1/scripts/training_setup_dev1.sh
bash: DEV1/scripts/training_setup_dev1.sh: No such file or directory
[training@localhost ~]$ clear

[training@localhost ~]$ $DEV1/scripts/training_setup_dev1.sh
[Dev] Setting up for Dev1Training
* Setting up Eclipse workspace for Dev1
* /home/training/workspace already exists, moving to /home/training/workspace.save.dev1
* ERROR: /home/training/workspace.save.dev1 already exists. Delete it and retry.
* Enabling and starting services required for Dev1 Training
Started Impala Catalog Server (catalogd) : [ OK ]
Stopping Apache Archiva...
Apache Archiva was not running.
Stopped Hive Server2: [ OK ]
* Disabling services not required for Dev1 Training
no secondarynamenode to stop
Stopped Hadoop secondarynamenode: [ OK ]
Flume agent is not running [ OK ]
Stopping Hadoop HBase regionserver daemon: no regionserver to stop because no pid file /var/run/hbase/hbase-hbase-regionserver.pid
hbase-regionserver.
no master to stop because no pid file /var/run/hbase/hbase-hbase-master.pid
Stopped HBase master daemon: [ OK ]
no rest to stop because no pid file /var/run/hbase/hbase-hbase-rest.pid
Stopped HBase rest daemon: [ OK ]
no thrift to stop because no pid file /var/run/hbase/hbase-hbase-thrift.pid
Stopped HBase thrift daemon: [ OK ]
Stopped Kafka Server: [ OK ]
Solr server daemon is not running [ OK ]
[Dev] Done setting up for Dev1 Training
[training@localhost ~]$
```

- 2) See the contents of the root directory in HDFS.

```
[training@localhost ~]$ hdfs dfs -ls /
Found 5 items
drwxrwxrwx  - hbase  supergroup          0 2015-08-27 05:37 /hbase
drwxrwxrwx  - solr   supergroup          0 2015-08-27 05:32 /solr
drwxrwxrwt  - hdfs   supergroup          0 2015-08-27 05:38 /tmp
drwxrwxrwx  - hdfs   supergroup          0 2015-08-27 05:32 /user
drwxrwxrwx  - hdfs   supergroup          0 2015-08-27 05:31 /var
[training@localhost ~]$
```

- 3) See the contents of the /user directory.

```
[training@localhost ~]$ hdfs dfs -ls /user
Found 5 items
drwxrwxrwt  - mapred  hadoop           0 2015-08-27 05:31 /user/history
drwxrwxrwx  - hive    hive             0 2015-08-27 05:31 /user/hive
drwxrwxrwx  - oozie   oozie            0 2015-08-27 05:32 /user/oozie
drwxr-xr-x  - spark   spark            0 2015-08-27 05:32 /user/spark
drwxrwxrwx  - training supergroup      0 2015-08-27 05:31 /user/training
[training@localhost ~]$
```

- 4) See contents of the home directory.

```
[training@localhost ~]$ hdfs dfs -ls /user/training
[training@localhost ~]$
```

- 5) Creating a new directory for assignments called loudacre.

- 6) Change directory to local file system directory which contains the sample data.

#### Output for 5 & 6

```
[training@localhost ~]$
[training@localhost ~]$ hdfs dfs -ls /user/training
[training@localhost ~]$
[training@localhost ~]$ hdfs dfs -mkdir /loudacre
[training@localhost ~]$ cd $DEV1DATA
[training@localhost data]$ ls
activations  calllogs  devicestatus.bin  devicestatus.txt  kb  mostactivestations.sql  static_data  synonyms.txt  weblogs
base_stations.csv  chatlogs  devicestatus_small.txt  frostroad.txt  loudacre.sql  purplecow.txt  stopwords.txt  targetmodels.txt
```

- 7) Insert this directory into HDFS.

- 8) List the contents of the new HDFS directory.

#### Output for 7 & 8

```
[training@localhost data]$ hdfs dfs -put kb /loudacre/
[training@localhost data]$ hdfs dfs -ls /loudacre/kb/
Found 300 items
-rw-rw-rw- 1 training supergroup 1626 2019-01-31 18:42 /loudacre/kb/KBDOC-00001.html
-rw-rw-rw- 1 training supergroup 1298 2019-01-31 18:42 /loudacre/kb/KBDOC-00002.html
-rw-rw-rw- 1 training supergroup 1183 2019-01-31 18:42 /loudacre/kb/KBDOC-00003.html
-rw-rw-rw- 1 training supergroup 1029 2019-01-31 18:42 /loudacre/kb/KBDOC-00004.html
-rw-rw-rw- 1 training supergroup 1155 2019-01-31 18:42 /loudacre/kb/KBDOC-00005.html
-rw-rw-rw- 1 training supergroup 1109 2019-01-31 18:42 /loudacre/kb/KBDOC-00006.html
-rw-rw-rw- 1 training supergroup 1648 2019-01-31 18:42 /loudacre/kb/KBDOC-00007.html
-rw-rw-rw- 1 training supergroup 1303 2019-01-31 18:42 /loudacre/kb/KBDOC-00008.html
-rw-rw-rw- 1 training supergroup 1173 2019-01-31 18:42 /loudacre/kb/KBDOC-00009.html
-rw-rw-rw- 1 training supergroup 1017 2019-01-31 18:42 /loudacre/kb/KBDOC-00010.html
-rw-rw-rw- 1 training supergroup 1149 2019-01-31 18:42 /loudacre/kb/KBDOC-00011.html
-rw-rw-rw- 1 training supergroup 1117 2019-01-31 18:42 /loudacre/kb/KBDOC-00012.html
-rw-rw-rw- 1 training supergroup 1638 2019-01-31 18:42 /loudacre/kb/KBDOC-00013.html
-rw-rw-rw- 1 training supergroup 1282 2019-01-31 18:42 /loudacre/kb/KBDOC-00014.html
-rw-rw-rw- 1 training supergroup 1166 2019-01-31 18:42 /loudacre/kb/KBDOC-00015.html
-rw-rw-rw- 1 training supergroup 1017 2019-01-31 18:42 /loudacre/kb/KBDOC-00016.html
-rw-rw-rw- 1 training supergroup 1160 2019-01-31 18:42 /loudacre/kb/KBDOC-00017.html
-rw-rw-rw- 1 training supergroup 1105 2019-01-31 18:42 /loudacre/kb/KBDOC-00018.html
-rw-rw-rw- 1 training supergroup 1612 2019-01-31 18:42 /loudacre/kb/KBDOC-00019.html
-rw-rw-rw- 1 training supergroup 1274 2019-01-31 18:42 /loudacre/kb/KBDOC-00020.html
-rw-rw-rw- 1 training supergroup 1176 2019-01-31 18:42 /loudacre/kb/KBDOC-00021.html
-rw-rw-rw- 1 training supergroup 1024 2019-01-31 18:42 /loudacre/kb/KBDOC-00022.html
-rw-rw-rw- 1 training supergroup 1149 2019-01-31 18:42 /loudacre/kb/KBDOC-00023.html
-rw-rw-rw- 1 training supergroup 1103 2019-01-31 18:42 /loudacre/kb/KBDOC-00024.html
-rw-rw-rw- 1 training supergroup 1580 2019-01-31 18:42 /loudacre/kb/KBDOC-00025.html
-rw-rw-rw- 1 training supergroup 1242 2019-01-31 18:42 /loudacre/kb/KBDOC-00026.html
-rw-rw-rw- 1 training supergroup 1143 2019-01-31 18:42 /loudacre/kb/KBDOC-00027.html
-rw-rw-rw- 1 training supergroup 1013 2019-01-31 18:42 /loudacre/kb/KBDOC-00028.html
-rw-rw-rw- 1 training supergroup 1130 2019-01-31 18:42 /loudacre/kb/KBDOC-00029.html
-rw-rw-rw- 1 training supergroup 1101 2019-01-31 18:42 /loudacre/kb/KBDOC-00030.html
-rw-rw-rw- 1 training supergroup 1630 2019-01-31 18:42 /loudacre/kb/KBDOC-00031.html
-rw-rw-rw- 1 training supergroup 1258 2019-01-31 18:42 /loudacre/kb/KBDOC-00032.html
-rw-rw-rw- 1 training supergroup 1143 2019-01-31 18:42 /loudacre/kb/KBDOC-00033.html
-rw-rw-rw- 1 training supergroup 1003 2019-01-31 18:42 /loudacre/kb/KBDOC-00034.html
```

```
[training@localhost:~/t...]
```

- 9) Uploading directory and removing it.
- 10) Displaying the last 20 lines of the article.

### Output for 9 & 10

```
[training@localhost data]$ hdfs dfs -put $DEV1DATA/callogs /loudacre/  
[training@localhost data]$ hdfs dfs -rm -r /loudacre/callogs/  
19/01/31 19:00:49 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion  
Deleted /loudacre/callogs  
[training@localhost data]$ hdfs dfs -cat /loudacre/kb/KBD0C-00289.html | tail -n 20  
</ol>  
<h3>Steps to change the notification ringtone for texts on Ronin S2</h3>  
<ol>  
<li>From the Personalization screen, tap Settings.</li>  
<li>Tap Sounds.</li>  
<li>Tap Text Tone.</li>  
<li>Tap your preferred ringtone.</li>  
<li>You've now changed the ringtone.</li>  
</ol>  
<h3>Steps to change the notification ringtone for emails on Ronin S2</h3>  
<ol>  
<li>From the Personalization screen, tap Settings.</li>  
<li>Tap Sounds.</li>  
<li>Scroll to and tap New Mail.</li>  
<li>Tap your preferred ringtone.</li>  
<li>You've now changed the ringtone.</li>  
</ol>  
<h4>Tags: smartphone, changes, contact, themselves, yours</h4>  
</body>  
</html>[training@localhost data]$ █
```

Current workspace: "Workspace 1"

- 11) Retrieve local copy of file or directory from HDFS and view data copied into HDFS, then get data to from HDFS to local directory as 'article.html' and read it.

```
[training@localhost data]$ hdfs dfs -get /loudacre/kb/KBDOC-00289.html ~/article.html
[training@localhost data]$ █
[training@localhost ~]$ less ~/article.html █
```

---

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
 "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xml:lang="en" lang="en">
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
<meta name="device" content="Ronin S2" />
<meta name="docid" content="KBDOC-00289" />
<meta name="createddate" content="2012-07-21" />
<meta name="updateddate" content="2012-08-03" />
<title>Ronin S2 - Change the phone ringtone and notification sound</title>
<link rel="stylesheet" type="text/css" href="theme.css" />
</head>
<body>
<h1>Ronin S2 - Change the phone ringtone and notification sound</h1>
<h2>Document Id: KBDOC-00289</h2>
<h2>Created: 2012-07-21</h2>
<h2>Updated: 2012-08-03</h2>
<h3>Steps to change the phone ringtone for calls on your Ronin S2</h3>
<ol>
<li>From the Personalization screen, tap Settings.</li>
<li>Tap Sounds.</li>
<li>Tap Ringtone.</li>
<li>Tap your preferred ringtone.</li>
<li>You've now changed the ringtone.</li>
</ol>
<h3>Steps to change the notification ringtone for texts on Ronin S2</h3>
<ol>
<li>From the Personalization screen, tap Settings.</li>
<li>Tap Sounds.</li>
<li>Tap Text Tone.</li>
<li>Tap your preferred ringtone.</li>
<li>You've now changed the ringtone.</li>
</ol>
<h3>Steps to change the notification ringtone for emails on Ronin S2</h3>
<ol>
:█
```

- 12) Checking all the file system commands provided by HDFS.

```
[training@localhost data]$ hdfs dfs [generic options]
Usage: hadoop fs [generic options]
  [-appendToFile <localsrc> ... <dst>]
  [-cat [-ignoreCrc] <src> ...]
  [-checksum <src> ...]
  [-chgrp [-R] GROUP PATH...]
  [-chmod [-R] <MODE[,MODE]... | OCTALMODE> PATH...]
  [-chown [-R] [OWNER][[:GROUP]] PATH...]
  [-copyFromLocal [-f] [-p] [-l] <localsrc> ... <dst>]
  [-copyToLocal [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
  [-count [-q] [-h] <path> ...]
  [-cp [-f] [-p | -p[topax]] <src> ... <dst>]
  [-createSnapshot <snapshotDir> [<snapshotName>]]
  [-deleteSnapshot <snapshotDir> [<snapshotName>]]
  [-df [-h] [<path> ...]]
  [-du [-s] [-h] <path> ...]
  [-expunge]
  [-get [-p] [-ignoreCrc] [-crc] <src> ... <localdst>]
  [-getfacl [-R] <path>]
  [-getfattr [-R] {-n name | -d} [-e en] <path>]
  [-getmerge [-nl] <src> <localdst>]
  [-help [cmd ...]]
  [-ls [-d] [-h] [-R] [<path> ...]]
  [-mkdir [-p] <path> ...]
  [-moveFromLocal <localsrc> ... <dst>]
  [-moveToLocal <src> <localdst>]
  [-mv <src> ... <dst>]
  [-put [-f] [-p] [-l] <localsrc> ... <dst>]
  [-renameSnapshot <snapshotDir> <oldName> <newName>]
  [-rm [-f] [-r|-R] [-skipTrash] <src> ...]
  [-rmdir [--ignore-fail-on-non-empty] <dir> ...]
  [-setfacl [-R] [{--b|-k} {-m|-x <acl_spec>} <path>]||[--set <acl_spec> <pa
th>]]
  [-setfattr {-n name [-v value] | -x name} <path>]
  [-setrep [-R] [-w] <rep> <path> ...]
  [-stat [format] <path> ...]
```

- 13) Started Firefox on the VM
  - 14) Accessed the local host to open HUE at port number 8888.
  - 15) Signup popped up the first time and login was created with training as user and password.
- Output for 13,14 & 15**

Quick Start Wizard - Hue™ 3.7.0 - The Hadoop UI

Step 1: Check Configuration Step 2: Examples Step 3: Users Step 4: Go!

Checking current configuration

Configuration files located in `/etc/hue/conf.empty`

Potential misconfiguration detected. Fix and restart Hue.

<code>down</code>	The Oozie server is not available
<code>Hive Editor</code>	The application won't work without a running HiveServer2.
<code>Oozie Editor/Dashboard</code>	The app won't work without a running Oozie server

Back Next

Hue and the Hue logo are trademarks of Cloudera, Inc.

Firefox automatically sends some data to Mozilla so that we can improve your experience. Choose What I Share

- 16) Select file browser.

File Browser

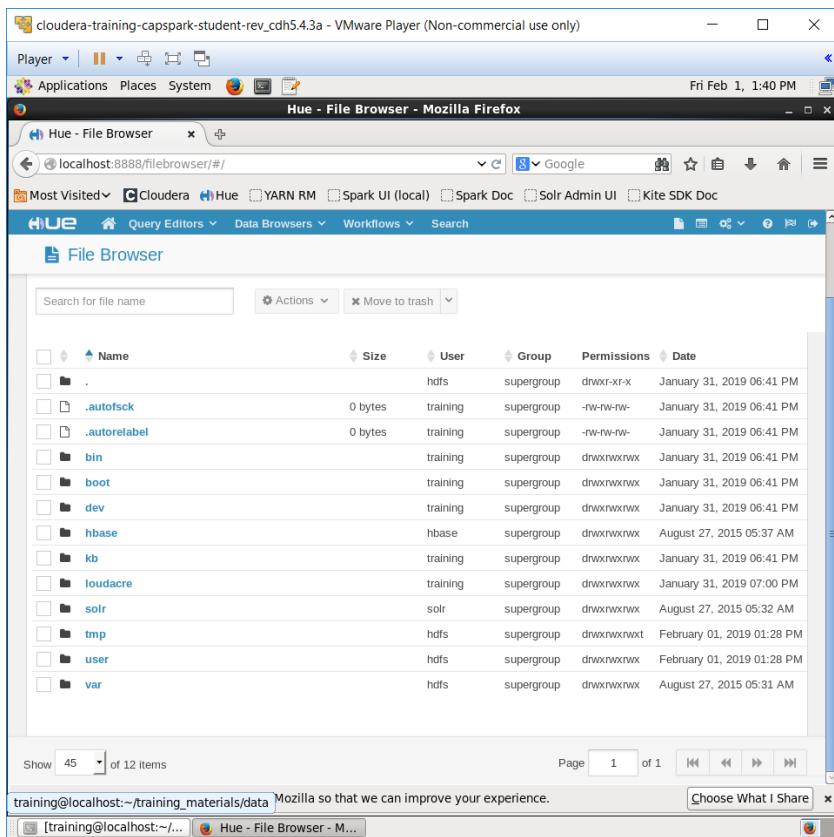
Search for file name Actions Move to trash

Home / user / training History Trash

Name	Size	User	Group	Permissions	Date
hdfs		hdfs	supergroup	drwxrwxrwx	February 01, 2019 01:28 PM
training		training	supergroup	drwxrwxrwx	August 27, 2015 05:31 AM

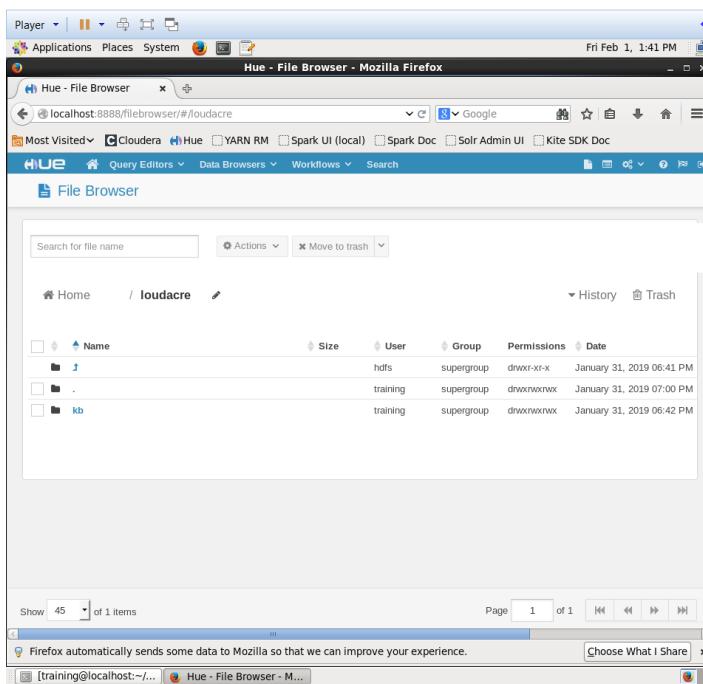
Show 45 of 0 items Page 1 of 1

17) On selection of / next to user, root directory is shown.



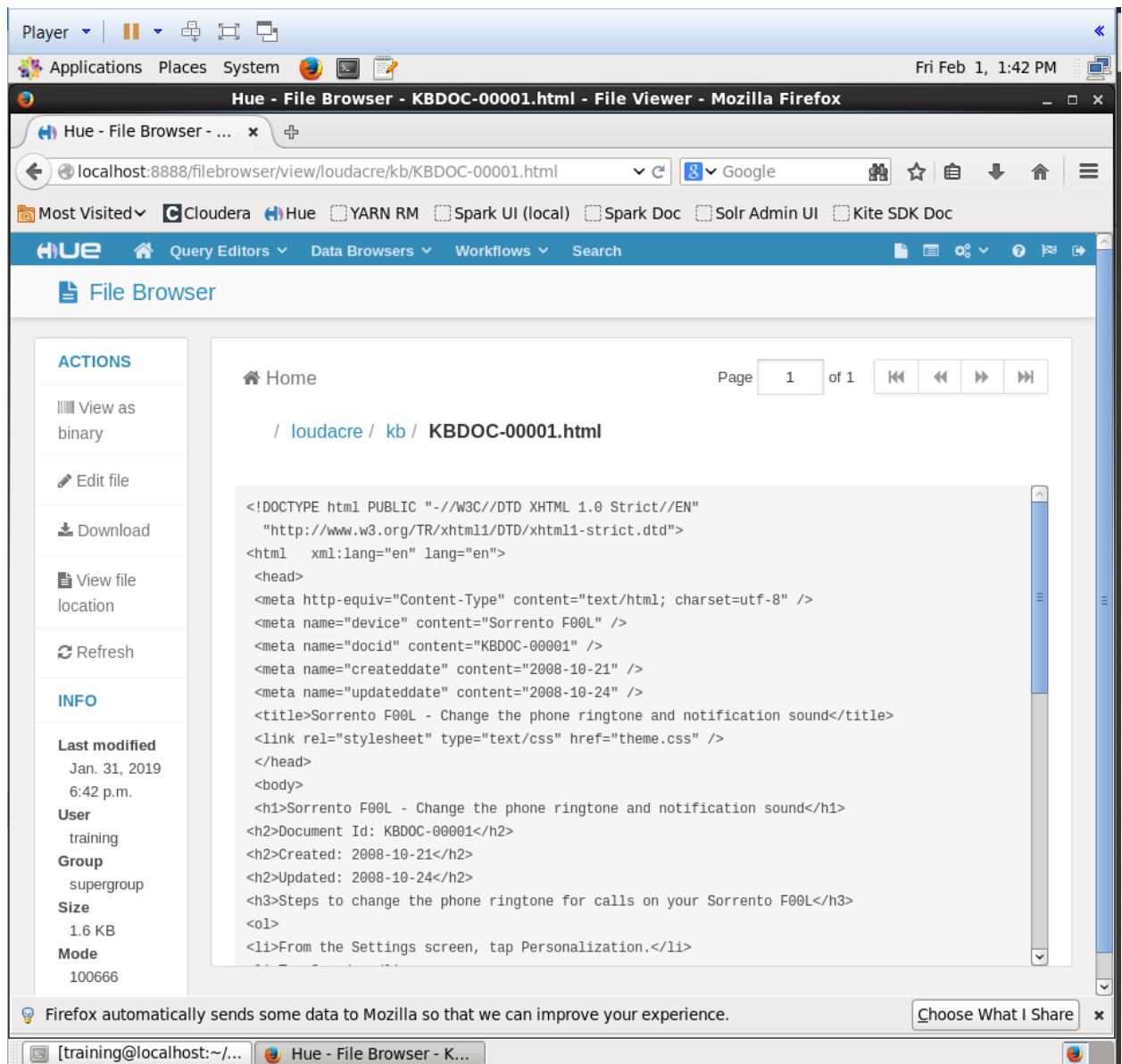
The screenshot shows a Mozilla Firefox browser window titled "Hue - File Browser". The address bar displays "localhost:8888/filebrowser/#/". The main content area is a table listing directory contents. The table has columns: Name, Size, User, Group, Permissions, and Date. The root directory "/" is listed first, followed by ".autofsck", ".autorelabel", "bin", "boot", "dev", "hbase", "kb", "loudacre", "solr", "tmp", "user", and "var". The "user" directory is selected, as indicated by a blue border around its row. The permissions for "user" are "drwxrwxrwx". The date for "user" is "February 01, 2019 01:28 PM". At the bottom of the browser window, there is a status message: "training@localhost:~/training\_materials/data Mozilla so that we can improve your experience." and a "Choose What I Share" button.

18) On selection of loudacre, its content is displayed.



The screenshot shows a Mozilla Firefox browser window titled "Hue - File Browser". The address bar displays "localhost:8888/filebrowser/#/loudacre". The main content area is a table listing directory contents. The table has columns: Name, Size, User, Group, Permissions, and Date. The "loudacre" directory is listed first, followed by "j", ".", and "kb". The "kb" directory is selected, as indicated by a blue border around its row. The permissions for "kb" are "drwxrwxrwx". The date for "kb" is "January 31, 2019 06:42 PM". At the bottom of the browser window, there is a status message: "Firefox automatically sends some data to Mozilla so that we can improve your experience." and a "Choose What I Share" button.

19) On selection of kb directory to view the articles uploaded.

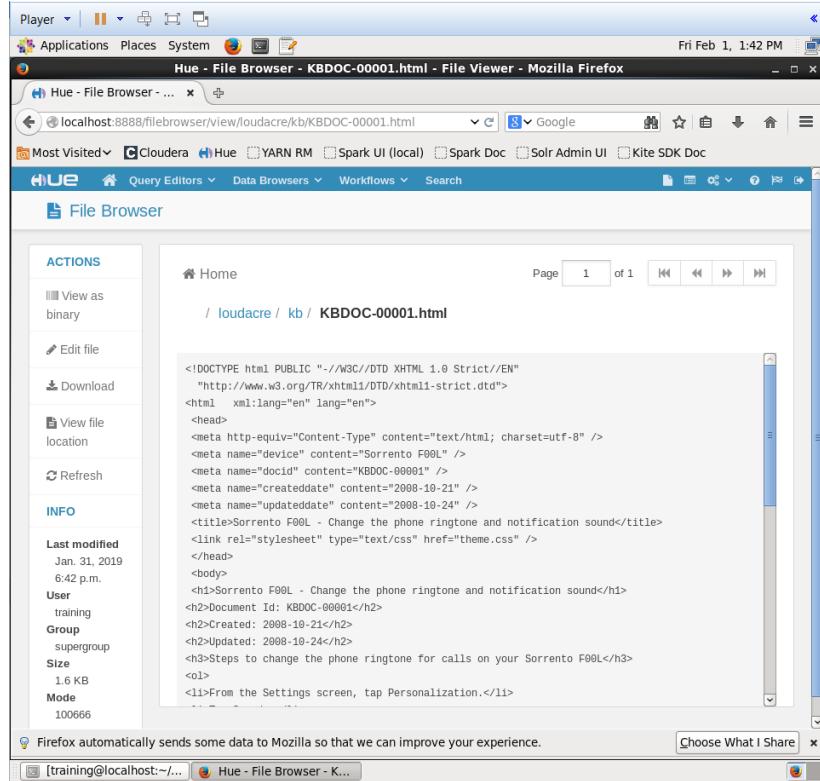


The screenshot shows a Mozilla Firefox window with the following details:

- Toolbar:** Player, Applications, Places, System, and a Firefox icon.
- Address Bar:** localhost:8888/filebrowser/view/loudacre/kb/KBDOC-00001.html - File Viewer - Mozilla Firefox. The date is Fri Feb 1, 1:42 PM.
- Page Content:** The main area shows the HTML code of the file KBDOC-00001.html. The code includes meta tags for XML and character encoding, and a title indicating it's about changing phone ringtone and notification sound for a Sorrento F00L device.
- Left Sidebar (Actions):**
  - ACTIONS**
    - View as binary
    - Edit file
    - Download
    - View file location
    - Refresh
  - INFO**
    - Last modified: Jan. 31, 2019, 6:42 p.m.
    - User: training
    - Group: supergroup
    - Size: 1.6 KB
    - Mode: 100666
- Bottom Status Bar:** Firefox automatically sends some data to Mozilla so that we can improve your experience. Choose What I Share.

- 20) On selection of KBDOC-00001.html its content is viewed.
- 21) Hue provides the page navigation function.
- 22) Return to directory by clicking View file Location
- 23) Click up arrow to return to loudacre.

### Output for 20, 21, 22 &23

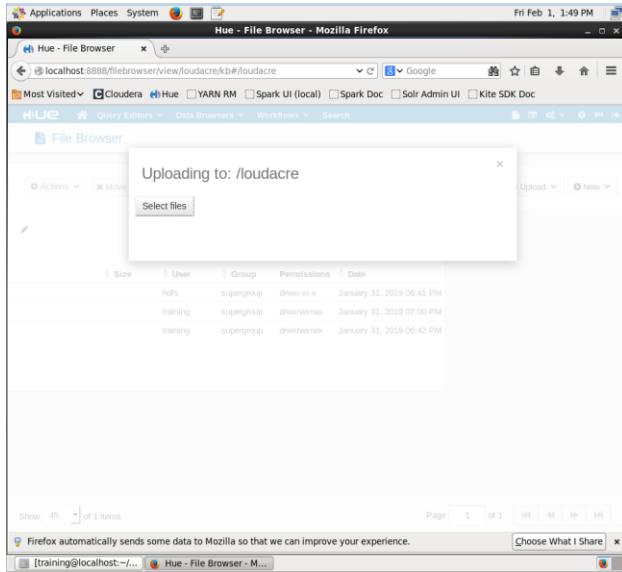


The screenshot shows a Mozilla Firefox window with the title "Hue - File Browser - KBDOC-00001.html - File Viewer - Mozilla Firefox". The address bar shows "localhost:8888/filebrowser/view/roudacre/kb/KBDOC-00001.html". The browser interface includes a toolbar, a menu bar, and a status bar indicating "Fri Feb 1, 1:42 PM". The main content area is a "File Browser" interface with a sidebar on the left containing "ACTIONS" and "INFO" sections. The "ACTIONS" section includes "View as binary", "Edit file", "Download", "View file location", and "Refresh". The "INFO" section shows "Last modified" (Jan. 31, 2019, 6:42 p.m.), "User" (training), "Group" (supergroup), "Size" (1.6 KB), "Mode" (100666). The central area displays the HTML content of KBDOC-00001.html, which includes meta-information like the document type, creation date, and update date, and a list of steps to change the phone ringtone.

File Browser					
<input type="text" value="Search for file name"/> <input type="button" value="Actions"/> <input type="button" value="Move to trash"/>					
<input type="button" value="Upload"/> <input type="button" value="New"/>					
<input type="button" value="Home"/> / roudacre / kb					
Name	Size	User	Group	Permissions	Date
KBDOC-00001.html	1.6 KB	training	supergroup	drwxrwxrwx	February 01, 2019 06:33 PM
KBDOC-00002.html	1.3 KB	training	supergroup	-rw-rw-rw-	January 31, 2019 06:42 PM
KBDOC-00003.html	1.2 KB	training	supergroup	-rw-rw-rw-	January 31, 2019 06:42 PM

File Browser					
<input type="text" value="Search for file name"/> <input type="button" value="Actions"/> <input type="button" value="Move to trash"/>					
<input type="button" value="Upload"/> <input type="button" value="New"/>					
<input type="button" value="Home"/> / roudacre					
Name	Size	User	Group	Permissions	Date
accounts	0 B	training	supergroup	drwxrwxrwx	February 01, 2019 06:33 PM
kb	0 B	training	supergroup	drwxrwxrwx	February 01, 2019 05:28 PM
webpage	0 B	training	supergroup	drwxrwxrwx	January 31, 2019 06:42 PM

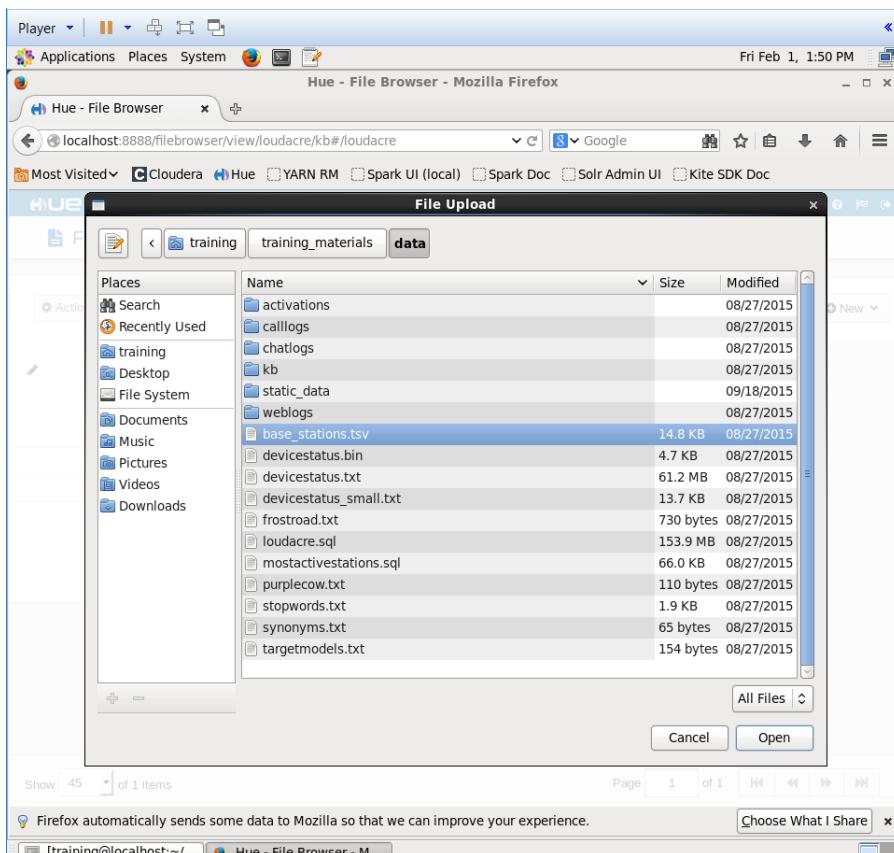
24) Click upload button to upload a file, then click select file to upload a particular file.



25) Browse to the directory of the file.

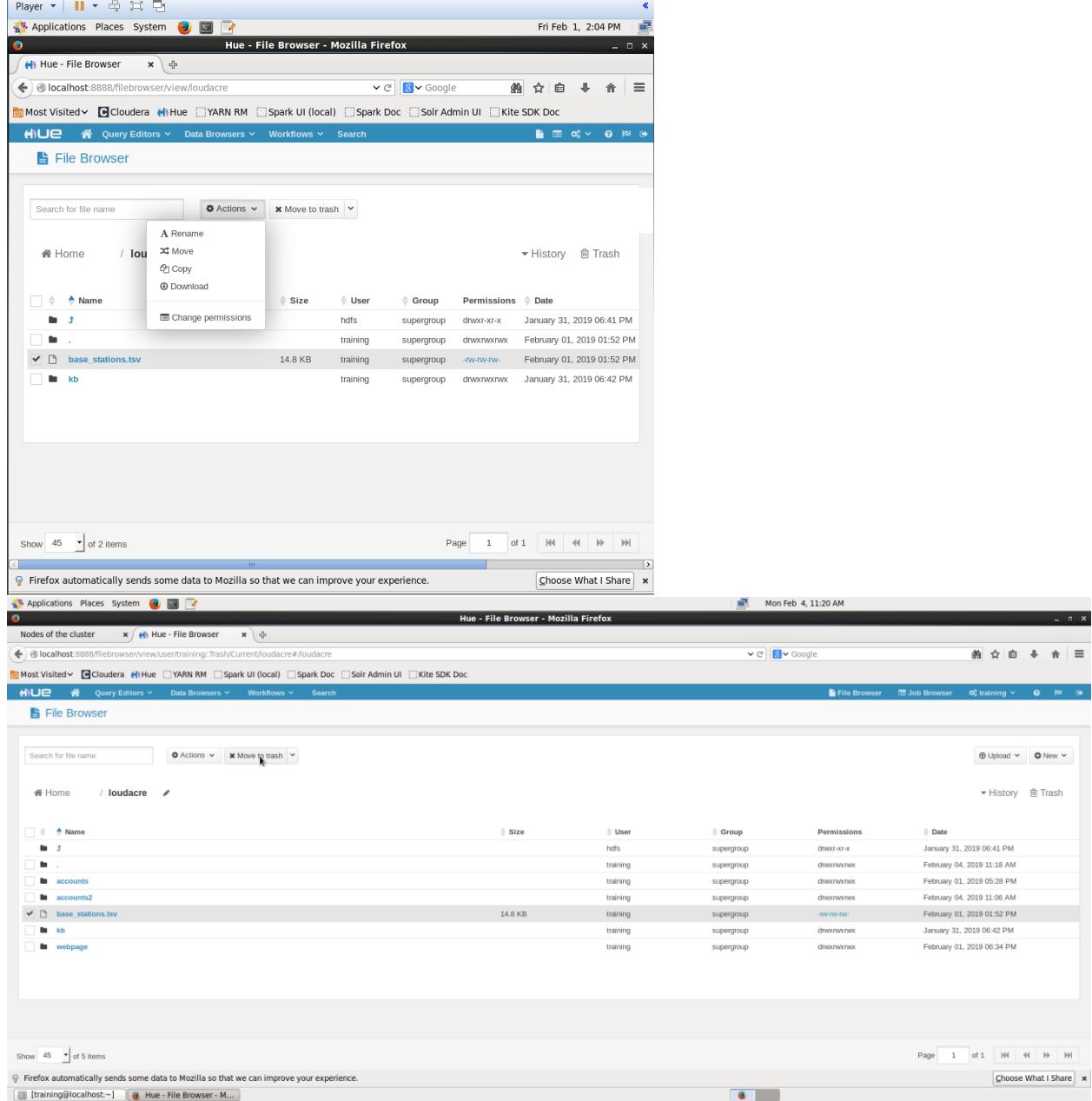
26) Choosing base\_station.tsv and click open.

**Output for 25 & 26**



- 27) Clicking the check box next to the file's icon and then clicking the Action button will display a list of actions that can be performed on the selected file.
- 28) Unneeded files were moved to trash by clicking 'Move to Trash" button

**Output for 27 & 28.**



The screenshot shows two instances of the Hue File Browser in Mozilla Firefox. The top instance is for the user 'loudacre' and the bottom instance is for the user 'training'. Both instances show a list of files in the directory '/loudacre'.

In the top instance (user 'loudacre'), a file named 'base\_stations.tsv' is selected. A context menu is open over this file, showing options: 'A Rename', 'Move', 'Copy', 'Download', and 'Change permissions'. Below the menu, the 'Actions' button is highlighted, and the 'Move to trash' button is being clicked.

In the bottom instance (user 'training'), the file 'base\_stations.tsv' is also selected. The 'Actions' button is highlighted, and the 'Move to trash' button is being clicked.

Both instances show a table of file details with columns: Name, Size, User, Group, Permissions, and Date. The file 'base\_stations.tsv' has a size of 14.8 KB, belongs to user 'training' and group 'supergroup', and has permissions '-rw-rw-rw-'. The date is February 01, 2019 01:52 PM.

## Yarn Job

- 1) Opening yarn resource manager UI in firefox through local host in port.
- 2) Taking note of values in the cluster metrics section.

**All Applications**

**Cluster Metrics**

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
0	0	0	0	0	0 B	2 GB	0 B	0	8	0	1	0	0	0	0

**User Metrics for dr.who**

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Containers Pending	Containers Reserved	Memory Used	Memory Pending	Memory Reserved	VCores Used	VCores Pending	VCores Reserved
0	0	0	0	0	0	0	0 B	0 B	0 B	0	0	0

Showing 0 to 0 of 0 entries

- 3) Clicking on the Nodes link to see list of worker nodes.
- 4) On click of Node HTTP Address Node Manager UI is displayed – we can view all the statistics of the node.

### Output for 3 & 4

**Nodes of the cluster**

**Cluster Metrics**

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
0	0	0	0	0	0 B	2 GB	0 B	0	8	0	1	0	0	0	0

**User Metrics for dr.who**

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Containers Pending	Containers Reserved	Memory Used	Memory Pending	Memory Reserved	VCores Used	VCores Pending	VCores Reserved
0	0	0	0	0	0	0	0 B	0 B	0 B	0	0	0

Showing 1 to 1 of 1 entries

**http://localhost:8042/node**

**NodeManager information**

Total Vmem allocated for Containers	4.20 GB
Vmem enforcement enabled	false
Total Pmem allocated for Container	2 GB
Pmem enforcement enabled	true
Total VCores allocated for Containers	8
NodeHealthStatus	true
LastNodeHealthTime	Fri Feb 01 15:34:30 PST 2019
NodeHealthReport	
Node Manager Version:	2.6.0-cdh5.4.3 from 4cd9f51a3f1ef748d45b8d77d0f211ad44296d4b by jenkins source checksum 8bacc650a3181bc0fb8e816dc81edd on 2015-06-25T02:41Z
Hadoop Version:	2.6.0-cdh5.4.3 from 4cd9f51a3f1ef748d45b8d77d0f211ad44296d4b by jenkins source checksum 4acea6ac185376e0b48b33695e88e7a7 on 2015-06-25T02:34Z

- 5) Return to Resource Manager on expanding Resource Manager and clicking RM Home.

- 6) Changing to the exercise directory.

```
[training@localhost data]$ cd $DEV1/exercises/yarn
[training@localhost yarn]$
```

- 7) Running the word counter py script on Yarn cluster, counting frequency of word.

```
[training@localhost data]$ cd $DEV1/exercises/yarn
[training@localhost yarn]$ spark-submit --master yarn-cluster wordcount.py /loud
acre/KB/*
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/lib/zookeeper/lib/slf4j-log4j12-1.7.5.jar!
/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/lib/flume-ng/lib/slf4j-log4j12-1.7.5.jar!
/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
19/02/01 15:42:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library
for your platform... using builtin-java classes where applicable
19/02/01 15:42:20 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0
:8032
19/02/01 15:42:21 INFO yarn.Client: Requesting a new application from cluster wi
th 1 NodeManagers
19/02/01 15:42:21 INFO yarn.Client: Verifying our application has not requested
more than the maximum memory capability of the cluster (1024 MB per container)
19/02/01 15:42:21 INFO yarn.Client: Will allocate AM container, with 896 MB memo
ry including 384 MB overhead
19/02/01 15:42:21 INFO yarn.Client: Setting up container launch context for our
```

8) Open hue in firefox

9) On selecting job browser we can see the current running and recently completed application.

10) The application also allow to click the application id to see the details of the running application

**Output for 8,9 & 10.**

The screenshot shows the Hue interface for managing Hadoop clusters. On the left, there's a sidebar with a 'hadoop' logo, navigation links for 'Cluster', 'Nodes', 'Applications', and 'Scheduler', and a 'Tools' section. The main content area is titled 'All Applications' and displays 'Cluster Metrics' and 'User Metrics for dr.who'. Below these are two tables: one for 'Cluster Metrics' and another for 'User Metrics'. The 'User Metrics' table shows a single entry for 'application\_1549055199617\_0001' with details: User: 'training', Name: 'wordcount.py', Application Type: 'SPARK', Queue: 'root.training', Start Time: 'Fri Feb 1 15:42:40 -0800 2019', State: 'RUNNING', Final Status: 'UNDEFINED', Running Containers: '2', Allocated CPU: '2', Allocated Memory MB: '2048', Progress: '0%', and Tracking UI: 'ApplicationMaster'. A search bar and a 'Show 20 entries' button are also present. At the bottom, a message says 'Showing 1 to 1 of 1 entries'.

The screenshot shows the Hue 'Job Browser' page. The top navigation bar includes links for 'Query Editors', 'Data Browsers', 'Workflows', and 'Search'. The main content area is titled 'Job Browser' and features a search bar with 'Username: training' and 'Text: Search for text'. Below this are four buttons: 'Succeeded' (green), 'Running' (orange), 'Failed' (red), and 'Killed' (dark grey). A table lists application logs with columns: Logs, ID, Name, Status, User, Maps, Reduces, Queue, Priority, Duration, and Submitted. Two entries are shown: one 'RUNNING' application '1549055199617\_0002' with status 'RUNNING', user 'training', and duration '29s'; and one 'SUCCEEDED' application '1549055199617\_0001' with status 'SUCCEEDED', user 'training', and duration '2m:5s'. A 'Kill' button is visible for the running application. At the bottom, a message says 'Showing 1 to 2 of 2 entries'.

11) On reloading yarn RM page in firefox, we can see the application which was started.

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU	Allocated Memory MB	Progress	Tracking UI
application_1549055199617_0003	training	wordcount.py	SPARK	root.training	Fri Feb 1 16:05:25 2019	N/A	RUNNING	UNDEFINED	2	2	2048	<div style="width: 100%;">ApplicationMaster</div>	<a href="#">History</a>
application_1549055199617_0002	training	wordcount.py	SPARK	root.training	Fri Feb 1 16:01:33 2019	Fri Feb 1 16:03:31 2019	FINISHED	SUCCEEDED	N/A	N/A	N/A	<div style="width: 100%;">History</div>	<a href="#">History</a>
application_1549055199617_0001	training	wordcount.py	SPARK	root.training	Fri Feb 1 15:42:24 2019	Fri Feb 1 15:44:29 2019	FINISHED	SUCCEEDED	N/A	N/A	N/A	<div style="width: 100%;">History</div>	<a href="#">History</a>

12) On selection of Nodes we can see the node details.

13) select Node HTTP Address and open Node Manager UI.

Output 12 & 13

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU	Allocated Memory MB	Progress	Tracking UI
application_1549055199617_0003	training	wordcount.py	SPARK	root.training	Fri Feb 1 16:05:25 2019	N/A	RUNNING	UNDEFINED	2	2	2048	<div style="width: 100%;">ApplicationMaster</div>	<a href="#">History</a>
application_1549055199617_0002	training	wordcount.py	SPARK	root.training	Fri Feb 1 16:01:33 2019	Fri Feb 1 16:03:31 2019	FINISHED	SUCCEEDED	N/A	N/A	N/A	<div style="width: 100%;">History</div>	<a href="#">History</a>
application_1549055199617_0001	training	wordcount.py	SPARK	root.training	Fri Feb 1 15:42:24 2019	Fri Feb 1 15:44:29 2019	FINISHED	SUCCEEDED	N/A	N/A	N/A	<div style="width: 100%;">History</div>	<a href="#">History</a>

Application Overview	
User:	training
Name:	wordcount.py
Application Type:	SPARK
Application Tags:	State: RUNNING
FinalStatus:	UNDEFINED
Started:	Mon Feb 04 00:07:13 -0800 2019
Elapsed:	42sec
Tracking URL:	ApplicationMaster
Diagnostics:	

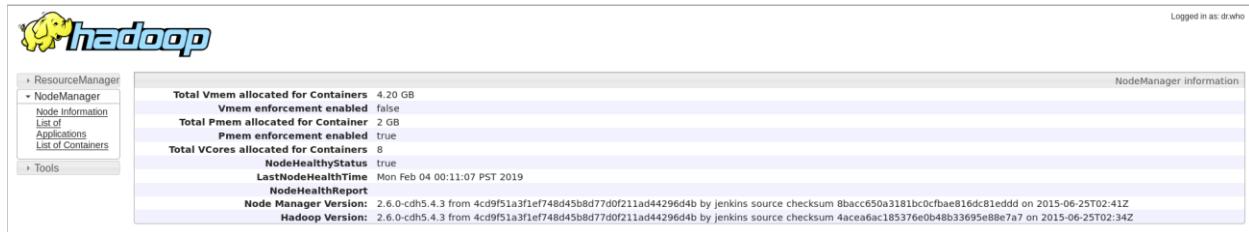
Application Metrics	
Total Resource Preempted:	<memory:0, vCores:0>
Total Number of Non-AM Containers Preempted:	0
Total Number of AM Containers Preempted:	0
Resource Preempted from Current Attempt:	<memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt:	0
Aggregate Resource Allocation:	66165 MB-seconds, 63 vcore-seconds

ApplicationMaster	
Attempt Number	1
Start Time	Mon Feb 04 00:07:13 -0800 2019
Node	localhost:8042
Logs	logs

14) While the application is running clicking on List of Applications will display the application submitted.

15) Clicking on List Containers will display the containers the RM has allocated for the current application.

### Output for 14 & 15



hadoop

Logged in as: dr.who

ResourceManager

NodeManager

Node Information

List of Applications

List of Containers

Tools

Total Vmem allocated for Containers: 4.20 GB  
Vmem enforcement enabled: false

Total Pmem allocated for Container: 2 GB  
Pmem enforcement enabled: true

Total VCores allocated for Containers: 8

NodeHealthyStatus: true

LastNodeHealthTime: Mon Feb 04 00:11:07 PST 2019

NodeHealthReport

Node Manager Version: 2.6.0-cdh5.4.3 from 4cd9f51a3f1ef748d45b8d77d0f211ad44296d4b by jenkins source checksum 8bacc650a3181bc0cfbae816dc81eddd on 2015-06-25T02:41Z

Hadoop Version: 2.6.0-cdh5.4.3 from 4cd9f51a3f1ef748d45b8d77d0f211ad44296d4b by jenkins source checksum 4acea6ac185376e0b48b33695e88e7a7 on 2015-06-25T02:34Z



hadoop

Logged in as: dr.who

ResourceManager

RM Home

NodeManager

Tools

All containers running on this node

Show: 20 entries

ContainerId	ContainerState	logs
container_1549055199617_0004_01_000001	RUNNING	logs
container_1549055199617_0004_01_000002	RUNNING	logs

Showing 1 to 2 of 2 entries

## MySQL to Sqoop

- 1) Opening Terminal Window.
- 2) Running sqoop help command to familiarize with options.

### Output for 1 and 2

```
[training@localhost yarn]$ sqoop help
19/02/01 16:26:47 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5-cdh5.4.3
usage: sqoop COMMAND [ARGS]
```

Available commands:

codegen	Generate code to interact with database records
create-hive-table	Import a table definition into Hive
eval	Evaluate a SQL statement and display the results
export	Export an HDFS directory to a database table
help	List available commands
import	Import a table from a database to HDFS
import-all-tables	Import tables from a database to HDFS
import-mainframe	Import datasets from a mainframe server to HDFS
job	Work with saved jobs
list-databases	List available databases on a server
list-tables	List available tables in a database
merge	Merge results of incremental imports
metastore	Run a standalone Sqoop metastore
version	Display version information

See 'sqoop help COMMAND' for information on a specific command.

### 3) Listing tables in loudacre

```
[training@localhost yarn]$ sqoop list-tables --connect jdbc:mysql://localhost/lu  
dacre --username training --password training  
19/02/01 16:32:26 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5-cdh5.4.3  
19/02/01 16:32:26 WARN tool.BaseSqoopTool: Setting your password on the command-  
line is insecure. Consider using -P instead.  
19/02/01 16:32:28 INFO manager.MySQLManager: Preparing to use a MySQL streaming  
resultset.  
accountdevice  
accounts  
basetations  
calllog  
customerservicerep  
device  
knowledgebase  
mostactivesstations  
toparticles  
webpage  
websitehit  
[training@localhost yarn]$ █
```

### 4) Running sqoop import command to see options.

```
[training@localhost yarn]$ sqoop import --help  
19/02/01 16:33:29 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5-cdh5.4.3  
usage: sqoop import [GENERIC-ARGS] [TOOL-ARGS]
```

#### Common arguments:

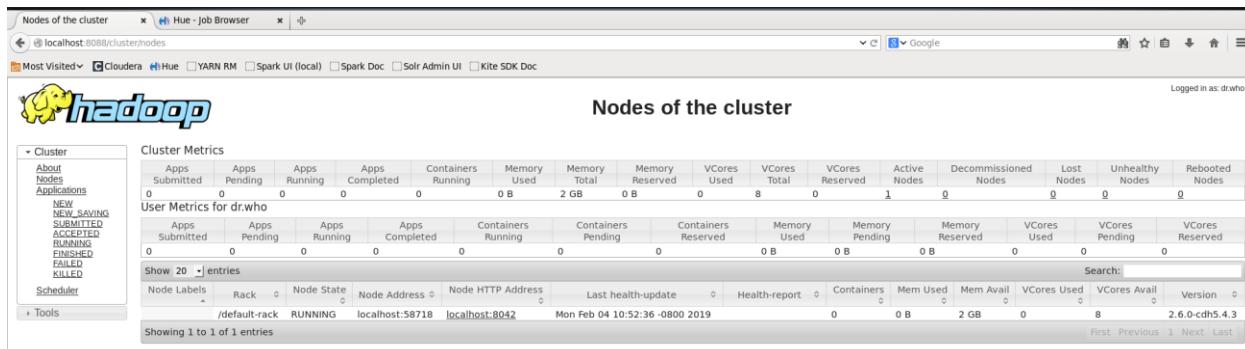
--connect <jdbc-uri>	Specify JDBC connect string
--connection-manager <class-name>	Specify connection manager class name
--connection-param-file <properties-file>	Specify connection parameters file
--driver <class-name>	Manually specify JDBC driver class to use
--hadoop-home <hdir>	Override \$HADOOP_MAPRED_HOME_ARG
--hadoop-mapred-home <dir>	Override \$HADOOP_MAPRED_HOME_ARG
--help	Print usage instructions
-P	Read password from console
--password <password>	Set authentication password
--password-alias <password-alias>	Credential provider password alias

5) importing accounts table in the loudacre database and saving it in HDFS under /loudacre.

6) (Optional) Sqoop job is can also be viewed in HUE or Yarn

## Output 5 & 6

```
[training@localhost yarn]$ sqoop import --connect jdbc:mysql://localhost/loudacre --username training --password training --table accounts --target-dir /loudacre/accounts --null-non-string '\\N'
19/02/01 16:45:09 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5-cdh5.4.3
19/02/01 16:45:09 WARN tool.BasesSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
19/02/01 16:45:10 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
19/02/01 16:45:10 INFO tool.CodeGenTool: Beginning code generation
19/02/01 16:45:12 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `accounts` AS t LIMIT 1
19/02/01 16:45:12 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `accounts` AS t LIMIT 1
19/02/01 16:45:12 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-training/compile/3fe3169e09d27580ece4606e8619fb17/accounts.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
19/02/01 16:45:23 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-training/compile/3fe3169e09d27580ece4606e8619fb17/accounts.jar
19/02/01 16:45:23 WARN manager.MySQLManager: It looks like you are importing from mysql.
19/02/01 16:45:23 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
19/02/01 16:45:23 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
19/02/01 16:45:23 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
19/02/01 16:45:23 INFO mapreduce.ImportJobBase: Beginning import of accounts
19/02/01 16:45:23 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
19/02/01 16:45:25 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
19/02/01 16:45:28 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
19/02/01 16:45:29 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
19/02/01 16:45:36 INFO db.DBInputFormat: Using read committed transaction isolation
19/02/01 16:45:36 INFO db.DataDrivenDBInputFormat: BoundingValsQuery: SELECT MIN(`acct_num`), MAX(`acct_num`) FROM `accounts`
19/02/01 16:45:36 INFO mapreduce.JobSubmitter: number of splits:4
19/02/01 16:45:37 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1549055199617_0005
19/02/01 16:45:38 INFO impl.YarnClientImpl: Submitted application application_1549055199617_0005
19/02/01 16:45:39 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1549055199617_0005
19/02/01 16:45:39 INFO mapreduce.Job: Running job: job_1549055199617_0005
19/02/01 16:46:03 INFO mapreduce.Job: Job job_1549055199617_0005 running in uber mode : false
19/02/01 16:46:04 INFO mapreduce.Job: map 0% reduce 0%
19/02/01 16:46:31 INFO mapreduce.Job: map 25% reduce 0%
19/02/01 16:46:53 INFO mapreduce.Job: map 50% reduce 0%
19/02/01 16:47:15 INFO mapreduce.Job: map 75% reduce 0%
```



Nodes of the cluster

Nodes of the cluster

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes
0	0	0	0	0	0 B	2 GB	0 B	0	8	0	1	0	0	0	0

User Metrics for dr.who

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Containers Pending	Containers Reserved	Memory Used	Memory Pending	Memory Reserved	VCores Used	VCores Pending	VCores Reserved
0	0	0	0	0	0	0	0 B	0 B	0 B	0	0	0

Show 20 - entries

Node Labels	Rack	Node State	Node Address	Node HTTP Address	Last health-update	Health-report	Containers	Mem Used	Mem Avail	VCores Used	VCores Avail	Version
/default-rack	RACK1	RUNNING	localhost:58718	localhost:8042	Mon Feb 04 10:52:36 -0800 2019	0	0 B	2 GB	0	8	2.6.0-cdh5.4.3	

Showing 1 to 1 of 1 entries

7) Listing contents of the accounts directory.

```
[training@localhost yarn]$ hdfs dfs -ls /loudacre/accounts
Found 5 items
-rw-rw-rw- 1 training supergroup          0 2019-02-01 16:47 /loudacre/accounts/_SU
CCESS
-rw-rw-rw- 1 training supergroup 4706617 2019-02-01 16:46 /loudacre/accounts/part-m-00000
-rw-rw-rw- 1 training supergroup 4693530 2019-02-01 16:46 /loudacre/accounts/part-m-00001
-rw-rw-rw- 1 training supergroup 4674529 2019-02-01 16:47 /loudacre/accounts/part-m-00002
-rw-rw-rw- 1 training supergroup 4662646 2019-02-01 16:47 /loudacre/accounts/part-m-00003
[training@localhost yarn]$ █
```

8) Display the last few part of the file using TAIL command.

9) The highest account Id is noted.

#### **Output of 8 &9**

```
[training@localhost yarn]$ hdfs dfs -tail /loudacre/accounts/part-m-00000
14-03-18 13:30:46.0
32434,2012-01-08 12:31:49.0,2013-12-25 04:55:12.0,Anne,Boyle,4215 McVaney Road,Reno,N
V,89535,7750546952,2014-03-18 13:30:46.0,2014-03-18 13:30:46.0
32435,2012-10-10 07:00:37.0,\N,Marjorie,Morrison,4417 Coffman Alley,Los Angeles,CA,90
042,2132809416,2014-03-18 13:30:46.0,2014-03-18 13:30:46.0
32436,2012-08-09 03:58:02.0,\N,Helen,Perez,1093 Columbia Road,Santa Barbara,CA,93490,
8054121301,2014-03-18 13:30:46.0,2014-03-18 13:30:46.0
32437,2012-01-14 10:27:37.0,\N,Melissa,Hayes,3434 Parkway Drive,Salem,OR,97394,503158
4725,2014-03-18 13:30:46.0,2014-03-18 13:30:46.0
32438,2012-08-22 20:40:31.0,\N,Violet,Searcy,2601 Twin Oaks Drive,San Francisco,CA,94
060,4152327199,2014-03-18 13:30:46.0,2014-03-18 13:30:46.0
32439,2012-12-15 08:45:17.0,2013-06-19 00:08:39.0,Eunice,Myers,4867 Reeves Street,Lon
g Beach,CA,90723,5628246641,2014-03-18 13:30:46.0,2014-03-18 13:30:46.0
32440,2012-09-19 09:11:15.0,\N,Robert,Huskey,953 Neville Street,San Jose,CA,95180,408
6716161,2014-03-18 13:30:46.0,2014-03-18 13:30:46.0
```

```
[training@localhost yarn]$ hdfs dfs -tail /loudacre/accounts/part-m-00001
:31:52.0
64874,2012-03-19 16:11:35.0,\N,Clara,Safford,4614 Spirit Drive,Glendale,CA,91217,7476
693718,2014-03-18 13:31:52.0,2014-03-18 13:31:52.0
64875,2012-10-10 20:07:25.0,\N,Erica,Oleary,3185 Fairway Drive,Long Beach,CA,90805,56
24213548,2014-03-18 13:31:52.0,2014-03-18 13:31:52.0
64876,2012-03-24 08:57:43.0,2014-01-27 09:14:34.0,Leslie,Jeanbaptiste,2573 Loving Acr
es Road,Kingman,AZ,86486,9286393819,2014-03-18 13:31:52.0,2014-03-18 13:31:52.0
64877,2012-04-24 05:52:44.0,\N,Jermaine,Harmon,4489 Shearwood Forest Drive,Burbank,CA
,91503,8186973079,2014-03-18 13:31:52.0,2014-03-18 13:31:52.0
64878,2012-08-04 01:49:37.0,\N,Mary,Hanson,522 Massachusetts Avenue,Oakland,CA,94676,
5108035413,2014-03-18 13:31:52.0,2014-03-18 13:31:52.0
64879,2012-10-21 23:36:04.0,\N,Richard,Alligood,4359 Agriculture Lane,San Diego,CA,92
087,6196073804,2014-03-18 13:31:52.0,2014-03-18 13:31:52.0
64880,2012-10-07 15:51:15.0,\N,Florence,Guevara,3507 School Street,San Francisco,CA,9
4183,4152020130,2014-03-18 13:31:52.0,2014-03-18 13:31:52.0
```

```
[training@localhost yarn]$ hdfs dfs -tail /loudacre/accounts/part-m-00002
3:32:51.0
97314,2013-01-18 14:05:59.0,2013-04-14 03:44:13.0,Eloy,Murphy,58 College View,San Mat
eo,CA,94493,6506652828,2014-03-18 13:32:51.0,2014-03-18 13:32:51.0
97315,2013-04-10 03:29:30.0,\N,Beatrice,Dodson,252 Farm Meadow Drive,Sacramento,CA,95
762,9165704077,2014-03-18 13:32:51.0,2014-03-18 13:32:51.0
97316,2013-03-26 14:13:25.0,\N,Gayle,Montelongo,364 Jail Drive,Kingman,AZ,86482,92855
24623,2014-03-18 13:32:51.0,2014-03-18 13:32:51.0
97317,2013-05-23 08:26:18.0,\N,Marc,Stock,1262 Ocala Street,Van Nuys,CA,91424,8184747
448,2014-03-18 13:32:51.0,2014-03-18 13:32:51.0
97318,2013-10-11 15:20:07.0,\N,Roger,Weyand,349 Concord Street,Tucson,AZ,85794,928539
1567,2014-03-18 13:32:51.0,2014-03-18 13:32:51.0
97319,2013-04-20 00:00:32.0,2014-01-31 00:41:43.0,Leona,Stannard,783 Stonecoal Road,P
alm Springs,CA,92240,7609402808,2014-03-18 13:32:51.0,2014-03-18 13:32:51.0
97320,2013-09-22 13:57:58.0,2014-03-14 15:55:49.0,Bryant,Nemec,4007 West Fork Drive,E
ugene,OR,97440,5411715739,2014-03-18 13:32:51.0,2014-03-18 13:32:51.0
```

```
[training@localhost yarn]$ hdfs dfs -tail /loudacre/accounts/part-m-00003
13:33:49.0
129755,2014-01-06 16:11:39.0,2014-02-21 11:22:56.0,John,Hale,1718 Irish Lane,San Bern
ardino,CA,92388,9097978236,2014-03-18 13:33:49.0,2014-03-18 13:33:49.0
129756,2014-01-26 19:34:06.0,\N,Danny,Graf,1104 Golden Ridge Road,San Jose,CA,95000,4
088662329,2014-03-18 13:33:49.0,2014-03-18 13:33:49.0
129757,2014-01-27 02:19:44.0,\N,John,McCall,4117 Conaway Street,Long Beach,CA,90609,5
624227213,2014-03-18 13:33:49.0,2014-03-18 13:33:49.0
129758,2014-03-14 20:25:15.0,\N,Robert,Pitt,2502 Heather Sees Way,Oakland,CA,94635,51
00736336,2014-03-18 13:33:49.0,2014-03-18 13:33:49.0
129759,2014-01-16 15:43:50.0,2014-02-19 07:31:53.0,Deborah,Hutchings,2071 Rodney Stre
et,Globe,AZ,85517,9288081544,2014-03-18 13:33:49.0,2014-03-18 13:33:49.0
129760,2014-01-06 04:24:56.0,\N,Zola,Tedder,1575 Linden Avenue,San Francisco,CA,94097
,4151620942,2014-03-18 13:33:49.0,2014-03-18 13:33:49.0
129761,2014-01-05 16:23:40.0,\N,Ruth,Ebersole,4684 Irving Place,Klamath Falls,OR,9764
6,5418511185,2014-03-18 13:33:49.0,2014-03-18 13:33:49.0
```

10) add\_new\_accounts.py script is run to add the latest accounts to MySql.

```
[training@localhost yarn]$ $DEV1/exercises/sqoop/add_new_accounts.py
Adding latest account data at 2019-02-01 17:23:33
Finished Adding Accounts
[training@localhost yarn]$ █
```

11) Using sqoop to import and append to accounts directory, on the last value on the acct\_num column largest account ID.

```
[training@localhost yarn]$ sqoop import --connect jdbc:mysql://localhost/loudacre --u
sername training --password training --incremental append --null-non-string '\\N' --t
able accounts --target-dir /loudacre/accounts --check-column acct_num --last-value 12
9761
19/02/01 17:26:35 INFO sqoop: Running Sqoop version: 1.4.5-cdh5.4.3
19/02/01 17:26:35 WARN tool.BaseSqoopTool: Setting your password on the command-line
is insecure. Consider using -P instead.
19/02/01 17:26:36 INFO manager.MySQLManager: Preparing to use a MySQL streaming resu
lset.
19/02/01 17:26:36 INFO tool.CodeGenTool: Beginning code generation
19/02/01 17:26:37 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `a
ccounts` AS t LIMIT 1
19/02/01 17:26:37 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `a
ccounts` AS t LIMIT 1
19/02/01 17:26:38 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-
mapreduce
Note: /tmp/sqoop-training/compile/df6492053f57c7762472bec624db9217/accounts.java uses
or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
19/02/01 17:26:46 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-training/
compile/df6492053f57c7762472bec624db9217/accounts.jar
19/02/01 17:26:50 INFO tool.ImportTool: Maximal id query for free form incremental im
port: SELECT MAX(`acct_num`) FROM `accounts`
19/02/01 17:26:50 INFO tool.ImportTool: Incremental import based on column `acct_num`
19/02/01 17:26:50 INFO tool.ImportTool: Lower bound value: 129761
19/02/01 17:26:50 INFO tool.ImportTool: Upper bound value: 129764
19/02/01 17:26:50 WARN manager.MySQLManager: It looks like you are importing from mys
ql.
19/02/01 17:26:50 WARN manager.MySQLManager: This transfer can be faster! Use the --d
irect
19/02/01 17:26:50 WARN manager.MySQLManager: option to exercise a MySQL-specific fast
path.
19/02/01 17:26:50 INFO manager.MySQLManager: Setting zero DATETIME behavior to conver
tToNull (mysql)
19/02/01 17:26:50 INFO mapreduce.ImportJobBase: Beginning import of accounts
```

12) Listing the contents of the accounts directory to verify sqoop imports.

```
[training@localhost yarn]$ hdfs dfs -ls /loudacre/accounts
Found 8 items
-rw-rw-rw- 1 training supergroup          0 2019-02-01 16:47 /loudacre/accounts/_SUCCESS
-rw-rw-rw- 1 training supergroup 4706617 2019-02-01 16:46 /loudacre/accounts/part-m-00000
-rw-rw-rw- 1 training supergroup 4693530 2019-02-01 16:46 /loudacre/accounts/part-m-00001
-rw-rw-rw- 1 training supergroup 4674529 2019-02-01 16:47 /loudacre/accounts/part-m-00002
-rw-rw-rw- 1 training supergroup 4662646 2019-02-01 16:47 /loudacre/accounts/part-m-00003
-rw-rw-rw- 1 training supergroup 129 2019-02-01 17:27 /loudacre/accounts/part-m-00004
-rw-rw-rw- 1 training supergroup 131 2019-02-01 17:27 /loudacre/accounts/part-m-00005
-rw-rw-rw- 1 training supergroup 131 2019-02-01 17:28 /loudacre/accounts/part-m-00006
-
```

13) Viewing entire contents of these files.

```
[training@localhost yarn]$ hdfs dfs -cat /loudacre/accounts/part-m-0000[456]
129762,2014-03-14 23:00:16.0,\N,Jesse,Anderson,123 Elm St,Reno,NV,89511,775-555-1212,
2019-02-01 17:23:33.0,2019-02-01 17:23:33.0
129763,2014-03-14 01:56:01.0,\N,Tom,Wheeler,125 Elm St,Palo Alto,CA,94301,650-555-151
5,2019-02-01 17:23:33.0,2019-02-01 17:23:33.0
129764,2014-03-14 06:15:26.0,\N,Ian,Wrigley,127 Elm St,Palo Alto,CA,94301,650-555-177
7,2019-02-01 17:23:33.0,2019-02-01 17:23:33.0
```

#### 14) Viewing few records in webpage table using sqoop eval.

```
nect jdbc:mysql://localhost/loudacre --username training --password training
19/02/01 18:29:50 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5-cdh5.4.3
19/02/01 18:29:50 ERROR tool.BaseSqoopTool: Error parsing arguments for eval:
19/02/01 18:29:50 ERROR tool.BaseSqoopTool: Unrecognized argument: --
19/02/01 18:29:50 ERROR tool.BaseSqoopTool: Unrecognized argument: query
19/02/01 18:29:50 ERROR tool.BaseSqoopTool: Unrecognized argument: select * FROM webpage
19/02/01 18:29:50 ERROR tool.BaseSqoopTool: Unrecognized argument: --connect
19/02/01 18:29:50 ERROR tool.BaseSqoopTool: Unrecognized argument: jdbc:mysql://localhost/loudacre
19/02/01 18:29:50 ERROR tool.BaseSqoopTool: Unrecognized argument: --username
19/02/01 18:29:50 ERROR tool.BaseSqoopTool: Unrecognized argument: training
19/02/01 18:29:50 ERROR tool.BaseSqoopTool: Unrecognized argument: --password
19/02/01 18:29:50 ERROR tool.BaseSqoopTool: Unrecognized argument: training

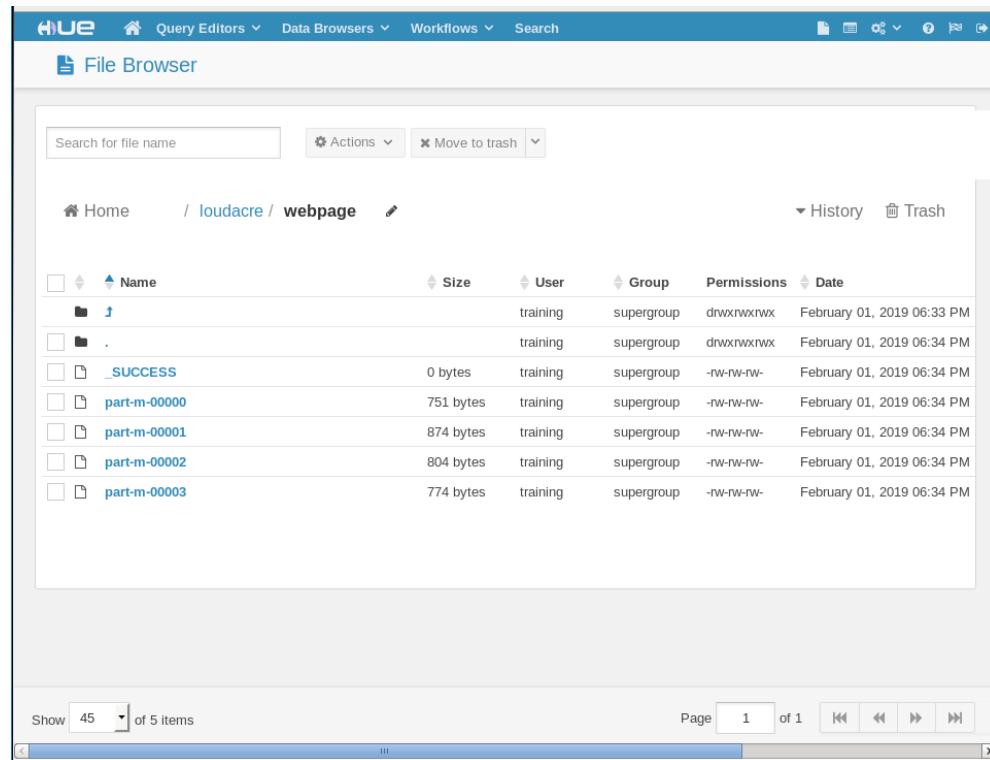
Try --help for usage instructions.
[training@localhost yarn]$ sqoop eval --query "select * FROM webpage LIMIT 10" --connect jdbc:mysql://localhost/loudacre --username training --password training
19/02/01 18:30:34 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5-cdh5.4.3
19/02/01 18:30:34 WARN tool.BaseSqoopTool: Setting your password on the command-line
is insecure. Consider using -P instead.
19/02/01 18:30:34 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.

-----
| web_page_num | web_page_file_name | associated_files |
-----
| 1            | sorrento_f00l_sales.html | theme.css,code.js,sorrento_f00l.jpg |
| 2            | titanic_2100_sales.html | theme.css,code.js,titanic_2100.jpg |
| 3            | meetoo_3.0_sales.html | theme.css,code.js,meetoo_3.0.jpg |
| 4            | meetoo_3.1_sales.html | theme.css,code.js,meetoo_3.1.jpg |
| 5            | ifruit_1_sales.html | theme.css,code.js,ifruit_1.jpg |
| 6            | ifruit_3_sales.html | theme.css,code.js,ifruit_3.jpg |
| 7            | ifruit_2_sales.html | theme.css,code.js,ifruit_2.jpg |
| 8            | ifruit_5_sales.html | theme.css,code.js,ifruit_5.jpg |
| 9            | titanic_1000_sales.html | theme.css,code.js,titanic_1000.jpg |
| 10           | meetoo_1.0_sales.html | theme.css,code.js,meetoo_1.0.jpg |
```

15) Using sqoop to import the webpage table, using tab character as a field terminator.

```
sername training --password training --table webpage --target-dir /loudacre/webpage --fields-terminated-by "\t"
19/02/01 18:33:43 INFO sqoop.Sqoop: Running Sqoop version: 1.4.5-cdh5.4.3
19/02/01 18:33:43 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
19/02/01 18:33:44 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
19/02/01 18:33:44 INFO tool.CodeGenTool: Beginning code generation
19/02/01 18:33:44 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `webpage` AS t LIMIT 1
19/02/01 18:33:44 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `webpage` AS t LIMIT 1
19/02/01 18:33:44 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/hadoop-mapreduce
Note: /tmp/sqoop-training/compile/d62bce2147fd1959d47657aab1e6a2b8/webpage.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
19/02/01 18:33:47 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-training/compile/d62bce2147fd1959d47657aab1e6a2b8/webpage.jar
19/02/01 18:33:47 WARN manager.MySQLManager: It looks like you are importing from mysql.
19/02/01 18:33:47 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
19/02/01 18:33:47 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
19/02/01 18:33:47 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
19/02/01 18:33:47 INFO mapreduce.ImportJobBase: Beginning import of webpage
19/02/01 18:33:47 INFO Configuration.deprecation: mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
19/02/01 18:33:48 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
19/02/01 18:33:48 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
19/02/01 18:33:49 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
19/02/01 18:33:50 INFO db.DBInputFormat: Using read committed transaction isolation
```

16) Using Hue to review the data files imported to /loudacre/webpage. Taking note of structure of the data.



The screenshot shows the Hue File Browser interface. The top navigation bar includes 'Query Editors', 'Data Browsers', 'Workflows', and 'Search'. Below the navigation is a 'File Browser' tab. The main area displays a file tree under the path '/loudacre / webpage'. The tree shows several files: a folder named '.\_SUCCESS', and four files named 'part-m-00000', 'part-m-00001', 'part-m-00002', and 'part-m-00003'. Each of these files is 0 bytes in size and was created by 'training' user on 'February 01, 2019 06:34 PM'. The 'Actions' dropdown menu is visible above the file list. At the bottom, there are buttons for 'History' and 'Trash', and a table showing the file details with columns: Name, Size, User, Group, Permissions, and Date.

Name	Size	User	Group	Permissions	Date
._SUCCESS	0 bytes	training	supergroup	-rw-rw-rw-	February 01, 2019 06:34 PM
part-m-00000	751 bytes	training	supergroup	-rw-rw-rw-	February 01, 2019 06:34 PM
part-m-00001	874 bytes	training	supergroup	-rw-rw-rw-	February 01, 2019 06:34 PM
part-m-00002	804 bytes	training	supergroup	-rw-rw-rw-	February 01, 2019 06:34 PM
part-m-00003	774 bytes	training	supergroup	-rw-rw-rw-	February 01, 2019 06:34 PM

Show 45 of 5 items

**ACTIONS**

- View as binary
- Edit file
- Download
- View file location
- Refresh

**INFO**

Last modified: Feb. 1, 2019 6:34 p.m.

User: training

Group: supergroup

Size: 751 bytes

Mode: 100666

File List:

- 1 sorrento\_f001\_sales.html theme.css,code.js,sorrento\_f001.jpg
- 2 titanic\_2100\_sales.html theme.css,code.js,titanic\_2100.jpg
- 3 meetoo\_3\_0\_sales.html theme.css,code.js,meetoo\_3.0.jpg
- 4 meetoo\_3\_1\_sales.html theme.css,code.js,meetoo\_3.1.jpg
- 5 ifruit\_1\_sales.html theme.css,code.js,ifruit\_1.jpg
- 6 ifruit\_3\_sales.html theme.css,code.js,ifruit\_3.jpg
- 7 ifruit\_2\_sales.html theme.css,code.js,ifruit\_2.jpg
- 8 ifruit\_5\_sales.html theme.css,code.js,ifruit\_5.jpg
- 9 titanic\_1000\_sales.html theme.css,code.js,titanic\_1000.jpg
- 10 meetoo\_1.0\_sales.html theme.css,code.js,meetoo\_1.0.jpg
- 11 sorrento\_f211\_sales.html theme.css,code.js,sorrento\_f211.jpg
- 12 ifruit\_4\_sales.html theme.css,code.js,ifruit\_4.jpg
- 13 sorrento\_f231\_sales.html theme.css,code.js,sorrento\_f231.jpg

Page 1 of 1

Choose What I Share

### Optional:

Using sqoop to import only accounts where the person lives in California and has an active account.

```
Try --help for usage instructions.
[training@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/loudacre
--username training --password training --table accounts --where "state='CA' and
acct_close_dt is NULL" --target-dir /loudacre/accounts2
```

---

```
19/02/04 11:03:22 INFO manager.SqlManager: Executing SQL statement: SELECT t.* F
ROM `accounts` AS t LIMIT 1
19/02/04 11:03:23 INFO manager.SqlManager: Executing SQL statement: SELECT t.* F
ROM `accounts` AS t LIMIT 1
19/02/04 11:03:22 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/lib/ha
oop-mapreduce
Note: /tmp/sqoop-training/compile/2c8218316da12637f774930215616ddc/accounts.java
uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
19/02/04 11:03:33 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-trai
ning/compile/2c8218316da12637f774930215616ddc/accounts.jar
19/02/04 11:03:34 WARN manager.MySQLManager: It looks like you are importing fro
m mysql.
19/02/04 11:03:34 WARN manager.MySQLManager: This transfer can be faster! Use th
e --direct
19/02/04 11:03:34 WARN manager.MySQLManager: option to exercise a MySQL-specific
fast path.
19/02/04 11:03:34 INFO manager.MySQLManager: Setting zero DATETIME behavior to c
onvertToNull (mysql)
19/02/04 11:03:34 INFO mapreduce.ImportJobBase: Beginning import of accounts
19/02/04 11:03:34 INFO Configuration.deprecation: mapred.job.tracker is deprecate
d. Instead, use mapreduce.jobtracker.address
19/02/04 11:03:35 INFO Configuration.deprecation: mapred.jar is deprecated. Inst
ead, use mapreduce.job.jar
19/02/04 11:03:39 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
19/02/04 11:03:39 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
19/02/04 11:03:47 INFO db.DBInputFormat: Using read committed transaction isolation
19/02/04 11:03:47 INFO db.DataDrivenDBInputFormat: BoundingValsQuery: SELECT MIN(`acct_num`), MAX(`acct_num`) FROM `accounts` WHERE ( state='CA' and acct_close_dt is NULL )
19/02/04 11:03:47 INFO mapreduce.JobSubmitter: number of splits:4
19/02/04 11:03:48 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1549306015403_0001
19/02/04 11:03:50 INFO impl.YarnClientImpl: Submitted application application_1549306015403_0001
19/02/04 11:03:51 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1549306015403_0001
19/02/04 11:03:51 INFO mapreduce.Job: Running job: job_1549306015403_0001
```

---

```
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=470
HDFS: Number of bytes written=8696802
HDFS: Number of read operations=16
HDFS: Number of large read operations=0
HDFS: Number of write operations=8
Job Counters
    Launched map tasks=4
    Other local map tasks=4
    Total time spent by all maps in occupied slots (ms)=0
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=88459
    Total vcore-seconds taken by all map tasks=88459
    Total megabyte-seconds taken by all map tasks=22645504
Map-Reduce Framework
    Map input records=61789
    Map output records=61789
    Input split bytes=470
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=1746
    CPU time spent (ms)=28460
    Physical memory (bytes) snapshot=473923584
    Virtual memory (bytes) snapshot=3383046144
    Total committed heap usage (bytes)=191889408
File Input Format Counters
    Bytes Read=0
File Output Format Counters
    Bytes Written=8696802
19/02/04 11:06:06 INFO mapreduce.ImportJobBase: Transferred 8.2939 MB in 147.2639 seconds (57.6718 KB/sec)
19/02/04 11:06:06 INFO mapreduce.ImportJobBase: Retrieved 61789 records.
```

## Review imported dataset using HDFS and HUE

```
[training@localhost ~]$ hdfs dfs -ls /loudacre/accounts2
Found 5 items
-rw-rw-rw- 1 training supergroup 0 2019-02-04 11:06 /loudacre/accounts2/_SUCCESS
-rw-rw-rw- 1 training supergroup 2294561 2019-02-04 11:04 /loudacre/accounts2/part-m-00000
-rw-rw-rw- 1 training supergroup 1997405 2019-02-04 11:05 /loudacre/accounts2/part-m-00001
-rw-rw-rw- 1 training supergroup 2094994 2019-02-04 11:05 /loudacre/accounts2/part-m-00002
-rw-rw-rw- 1 training supergroup 2309842 2019-02-04 11:06 /loudacre/accounts2/part-m-00003
```

Name	Size	User	Group	Permissions	Date
_SUCCESS	0 bytes	training	supergroup	drwxrwxrwx	February 04, 2019 11:04 AM
part-m-00000	2.2 MB	training	supergroup	drwxrwxrwx	February 04, 2019 11:06 AM
part-m-00001	1.9 MB	training	supergroup	drwxrwxrwx	February 04, 2019 11:04 AM
part-m-00002	2.0 MB	training	supergroup	drwxrwxrwx	February 04, 2019 11:05 AM
part-m-00003	2.2 MB	training	supergroup	drwxrwxrwx	February 04, 2019 11:06 AM

Timestamp	User	Log Data
2014-03-18 06:08:07,0	null	Duane,Ruiz,1823 Simpson Street,Pasadena,CA,91051,6262472463,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-03-15 06:51:05,0	null	Melissa,Carlson,3842 Paradise Lane,Torrance,CA,9052,4249518440,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-07-17 19:27:49,0	null	Ben,McFarlane,1574 Maloy Court,Sacramento,CA,94211,9168121036,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-03-31 13:26:14,0	null	Christine,Smith,2888 Ritter Avenue,San Diego,CA,92121,619772437,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-01-23 18:28:59,0	null	Johnny,Rodgers,9471 Joseph Street,Sacramento,CA,95734,9165124136,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-02-17 16:20:57,0	null	Rita,Rodriguez,1320 Spring Street,Sacramento,CA,95797,9163471136,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-03-18 03:56:18,0	null	John,White,2000 2nd Street,Long Beach,CA,90803,619772437,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-02-23 22:59:44,0	null	Wanda,Ferguson,Bell McAllister Street,Salinas,CA,93066,6310007172,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-01-23 04:09:27,0	null	Amelia,Santauria,4600 Coffman Alley,San Bernardino,CA,92370,9096113769,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-07-04 15:29:15,0	null	Erika,Moorhouse,3354 Cooks Mine Road,Alhambra,CA,91786,6367393772,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-11-28 20:24:17,0	null	Claudia,Allen,3673 Salice Way,Pasadena,CA,91132,6368168617,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-03-18 03:56:18,0	null	Heriberto,Johns,665 Morgan Street,North Hollywood,CA,91357,8188545227,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-04-11 19:47:34,0	null	Rene,Stripling,3235 Orphan Road,Santa Ana,CA,9734,057755704,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-10-21 19:34:44,0	null	Gail,Saunders,3385 Heavens Way,Mojave,CA,93510,6516253564,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-10-18 02:06:09,0	null	Erica,Todd,3852 Petunia Way,Long Beach,CA,90844,5629717794,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-09-09 02:21:29,0	null	Daniel,Eckert,1988 Mahler Street,Glendale,CA,91122,7473601607,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-12-21 08:58:42,0	null	Edward,Jones,4215 Ascrift Court,Long Beach,CA,90651,5628432260,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-05-15 19:35:05,0	null	James,Johnson,1827 Sycamore Lake Road,Marysville,CA,95920,530394997,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-06-29 03:51:53,0	null	Arthur,Brennan,2800 Cedar Street,San Mateo,CA,94407,6508042180,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-03-02 04:09:00,0	null	Robert,Willings,4702 10th Avenue,Long Beach,CA,90803,619772437,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-08-18 16:32:45,0	null	Allen,Jones,2943 Wakefield Street,McJaves,CA,93905,6616473825,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0
2012-01-25 07:55:13,0	null	James,Taylor,330 Vine Street,Redding,CA,96001,5309613375,2014-03-18 13:30:46,0,2014-03-18 13:30:46,0