

Running the WeatherTemperature example

Summary

This guide describes how to compile and run the WeatherTemperature program, a Java program implementing the map and reduce functions used to analyse the logs from the National Climatic Data Center (NCDC). The logs present the temperature information of a given year. The WeatherTemperature program analyse the data do present the highest recorded global temperature for each year in the dataset.

Compiling the WeatherTemperature program

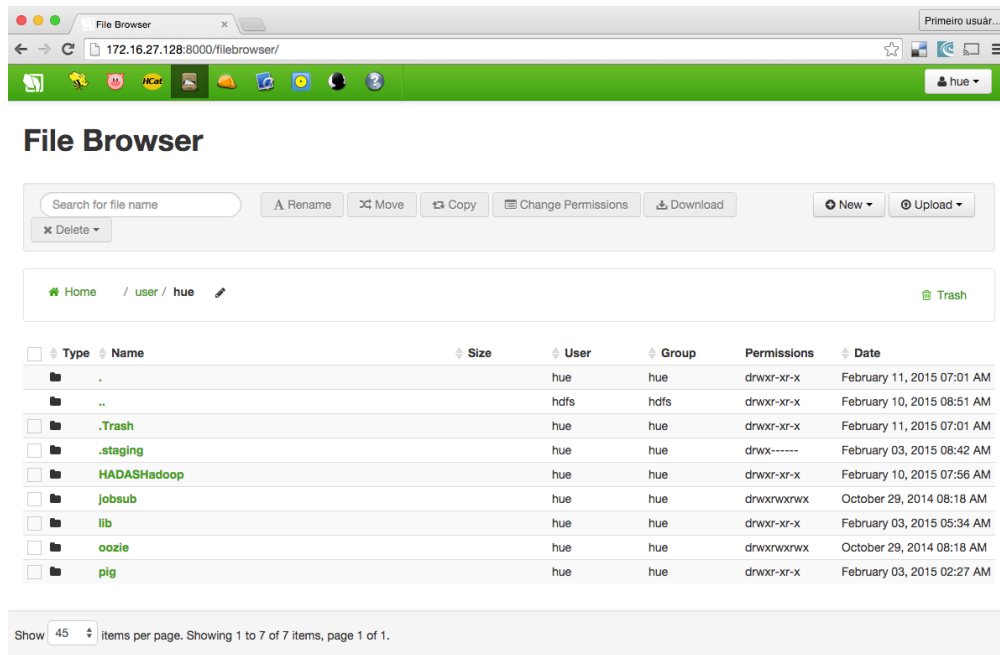
The WeatherTemperature program resides inside the WeatherTemperature folder. The folder is composed of the following files:

- `src/version1/MaxTemperatureMapper.java` - Contains the 'generical' map function implementation.
- `src/version1/MaxTemperatureReducer.java` - Contains the 'generical' reduce function implementation.
- `src/version2/MaxTemperatureMapper.java` - Contains the specific map function implementation (verify if the temperature is missing, if so, it applies a standard value).
- `src/version2/MaxTemperatureDriver.java` - Contains the execution main function of the program. This file define 'who' is the Mapper and the Reducer.
- **weather-temperature.jar** - The executable MaxTemperature file.
- `inputs/*` - The input files the be executed by the program.
- `src/lib/*`.jar - The hadoop library necessary to execute the program

Inside order to compile the WeatherTemperature program, you can execute it in your preferred Java Tool (like Eclipse), or execute it in the command line. The program is already compiled in the Java version 1.7 and compressed in the **weather-temperature.jar** file, ready to be executed.

Running the WeatherTemperature program in Hadoop - HortonWorks

First you may access the HDFS filebrowser (<http://172.16.27.128:8000/filebrowser/>) in create a folder for the examples, for example: **HADASHadoop**.

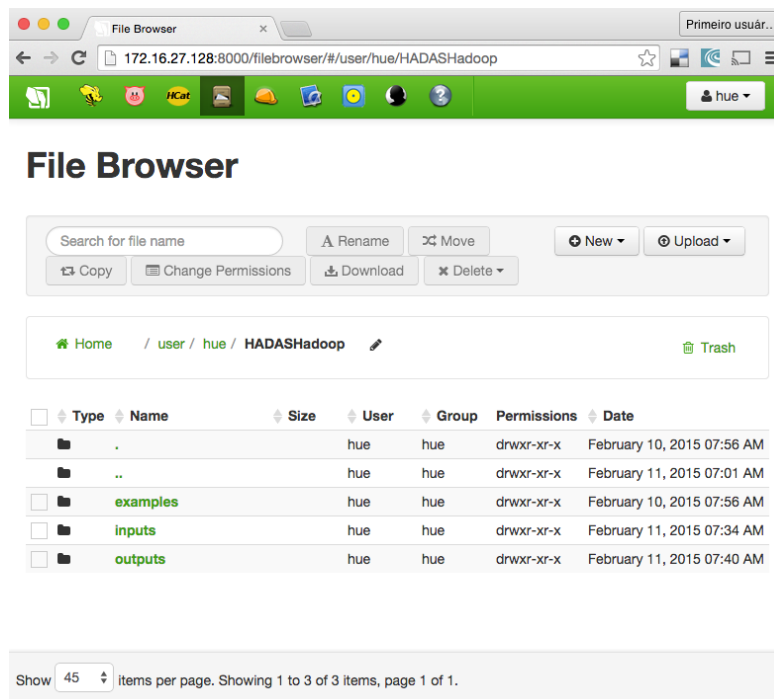


The screenshot shows the HDFS File Browser interface. The browser address bar displays `172.16.27.128:8000/filebrowser/`. The interface includes a search bar, navigation buttons (Rename, Move, Copy, Change Permissions, Download), and action buttons (New, Upload, Delete). The breadcrumb path is `Home / user / hue`. A table lists the contents of the directory:

Type	Name	Size	User	Group	Permissions	Date
Folder	.		hue	hue	drwxr-xr-x	February 11, 2015 07:01 AM
Folder	..		hdfs	hdfs	drwxr-xr-x	February 10, 2015 08:51 AM
Folder	.Trash		hue	hue	drwxr-xr-x	February 11, 2015 07:01 AM
Folder	.staging		hue	hue	drwx-----	February 03, 2015 08:42 AM
Folder	HADASHadoop		hue	hue	drwxr-xr-x	February 10, 2015 07:56 AM
Folder	jobsub		hue	hue	drwxrwxrwx	October 29, 2014 08:18 AM
Folder	lib		hue	hue	drwxr-xr-x	February 03, 2015 05:34 AM
Folder	oozie		hue	hue	drwxrwxrwx	October 29, 2014 08:18 AM
Folder	pig		hue	hue	drwxr-xr-x	February 03, 2015 02:27 AM

At the bottom, it shows "Show 45 items per page. Showing 1 to 7 of 7 items, page 1 of 1."

In this folder you may also create 3 new folders: **inputs** and **outputs**:



The screenshot shows the HDFS File Browser interface with the breadcrumb path `Home / user / hue / HADASHadoop`. A table lists the contents of the `HADASHadoop` directory:

Type	Name	Size	User	Group	Permissions	Date
Folder	.		hue	hue	drwxr-xr-x	February 10, 2015 07:56 AM
Folder	..		hue	hue	drwxr-xr-x	February 11, 2015 07:01 AM
Folder	examples		hue	hue	drwxr-xr-x	February 10, 2015 07:56 AM
Folder	inputs		hue	hue	drwxr-xr-x	February 11, 2015 07:34 AM
Folder	outputs		hue	hue	drwxr-xr-x	February 11, 2015 07:40 AM

At the bottom, it shows "Show 45 items per page. Showing 1 to 3 of 3 items, page 1 of 1."

In the **inputs** folder you may include all the input files. The outputs files will be generated after the hadoop execution.

```
placidoasouzaneto — root@sandbox:~ — bash — 80x17
neto:~ placidoasouzaneto$ scp /Users/placidoasouzaneto/TP-Hadoop/WeatherTemperature/weather-temperature.jar root@172.16.27.128:/usr/lib/
root@172.16.27.128's password:
weather-temperature.jar          100% 6228KB   6.1MB/s   00:00
neto:~ placidoasouzaneto$ scp /Users/placidoasouzaneto/TP-Hadoop/DP_Filtering-TopK/DP_Filtering-TopK/WeatherTemperature/
root@172.16.27.128's password:
.DS_Store          DP_Filtering-TopK.zip  WeatherTemperature.zip
DP_Filtering-TopK/ WeatherTemperature/
neto:~ placidoasouzaneto$ scp /Users/placidoasouzaneto/TP-Hadoop/DP_Filtering-TopK/filtering-topk.jar root@172.16.27.128:/usr/lib/
root@172.16.27.128's password:
filtering-topk.jar          100% 8713KB   8.5MB/s   00:01
neto:~ placidoasouzaneto$
```

To execute Hadoop In Hortonworks, it is necessary send the jar files to the server to be executed using the following ssh command. The root password for Hortonworks is **hadoop**.

After include the executable files in the server, you may login to execute the program, using the ssh command. The root password for Hortonworks is **hadoop**.

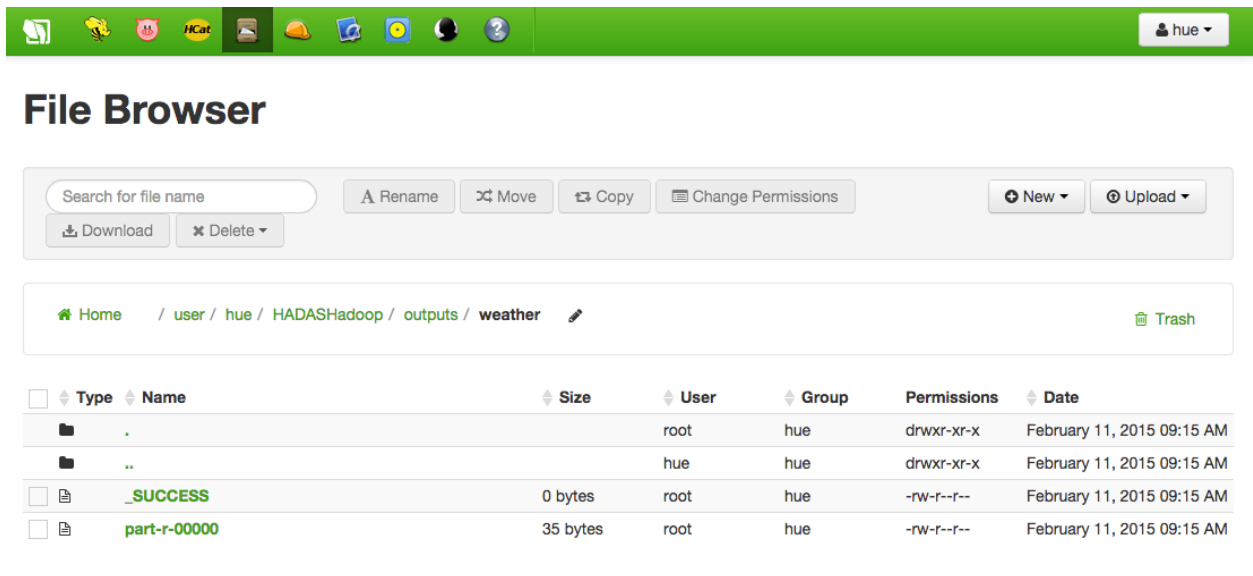
```
placidoasouzaneto — root@sandbox:~ — ssh — 80x17
neto:~ placidoasouzaneto$ ssh root@172.16.27.128
root@172.16.27.128's password:
Last login: Wed Feb 11 15:09:08 2015 from 172.16.27.1
[root@sandbox ~]#
```

Once Logged in Hadoop Server, you must execute the hadoop command:

`hadoop jar <ssh 'local' jar file> <hdfs input path> <hdfs output path>`

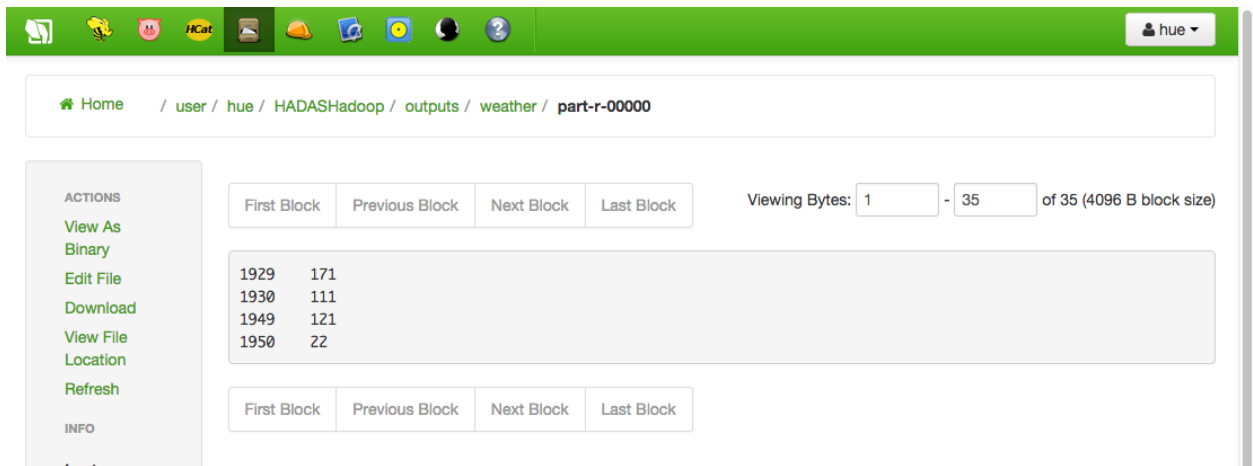
```
[root@sandbox ~]# hadoop jar /usr/lib/weather-temperature.jar /user/hue/HADASHadoop/inputs/sample2.txt /user/hue/HADASHadoop/outputs/weather
[root@sandbox ~]#
```

After that, the results can be reached in the HDFS file browser.



The screenshot shows the Hue File Browser interface. At the top, there is a green navigation bar with various icons and a user profile dropdown for 'hue'. Below this is the 'File Browser' title. A search bar and several action buttons (Rename, Move, Copy, Change Permissions, New, Upload, Download, Delete) are present. The breadcrumb path is 'Home / user / hue / HADASHadoop / outputs / weather'. A table lists the contents of the 'weather' directory:

Type	Name	Size	User	Group	Permissions	Date
Folder	.		root	hue	drwxr-xr-x	February 11, 2015 09:15 AM
Folder	..		hue	hue	drwxr-xr-x	February 11, 2015 09:15 AM
File	_SUCCESS	0 bytes	root	hue	-rw-r--r--	February 11, 2015 09:15 AM
File	part-r-00000	35 bytes	root	hue	-rw-r--r--	February 11, 2015 09:15 AM



The screenshot shows the Hue File Browser interface with the file 'part-r-00000' selected. The breadcrumb path is 'Home / user / hue / HADASHadoop / outputs / weather / part-r-00000'. On the left, there is a sidebar with 'ACTIONS' (View As Binary, Edit File, Download, View File Location, Refresh) and 'INFO'. The main content area shows a table of byte ranges and a 'Viewing Bytes' control.

Viewing Bytes: 1 - 35 of 35 (4096 B block size)

1929	171
1930	111
1949	121
1950	22