

Running the Design Pattern Filtering Top-K example

Summary

This guide describes how to compile and run the Filtering-TopK program, a Java program implementing the map and reduce functions. The point of this pattern is to find the best records for a specific criterion so that you can take a look at them and perhaps figure out what caused them to be so special. If you can define a ranking function or comparison function between two records that determines whether one is higher than the other, you can apply this pattern to use MapReduce to find the records with the highest value across your entire data set.

Applicability

This pattern requires a comparator function ability between two records. That is, we must be able to compare one record to another to determine which is “larger.”

The number of output records should be significantly fewer than the number of input records because at a certain point it just makes more sense to do a total ordering of the data set.

Compiling the Filtering-TopK program

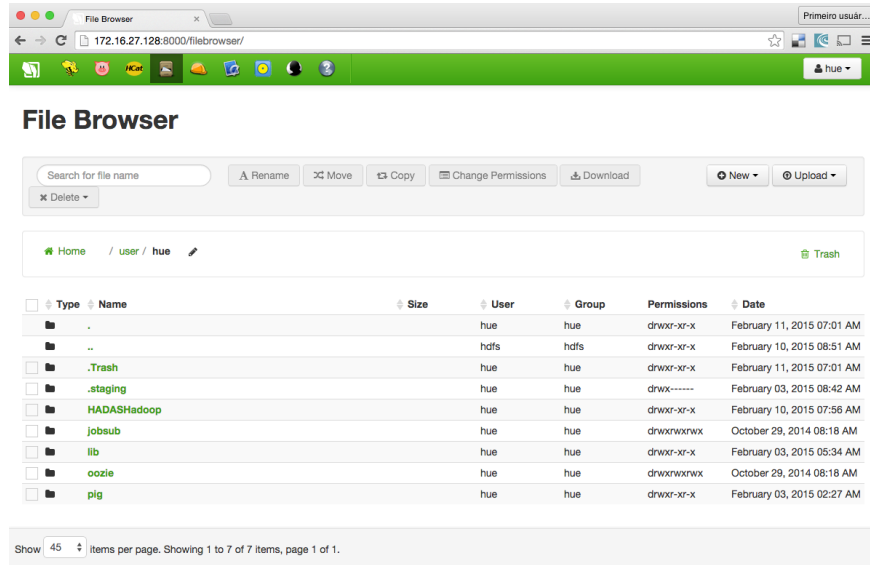
The Filtering-TopK program resides inside the DP_Filtering-TopK folder. The folder is composed of the following files:

- **src/TopTenDriver.java** - Contains the map and reduce function implementation. This file also contains the main function to execute the program
- **src/MRDPUtils.java** - This java file formats the .xml as input.
- **filtering-topk.jar** - The executable filtering-topk file.
- **inputs/*** - The input files to be executed by the program.
- **src/lib/*.jar** - The hadoop library necessary to execute the program

In order to compile the Filtering-TopK 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 **filtering-topk.jar** file, ready to be executed.

Running the Filtering-TopK 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**.

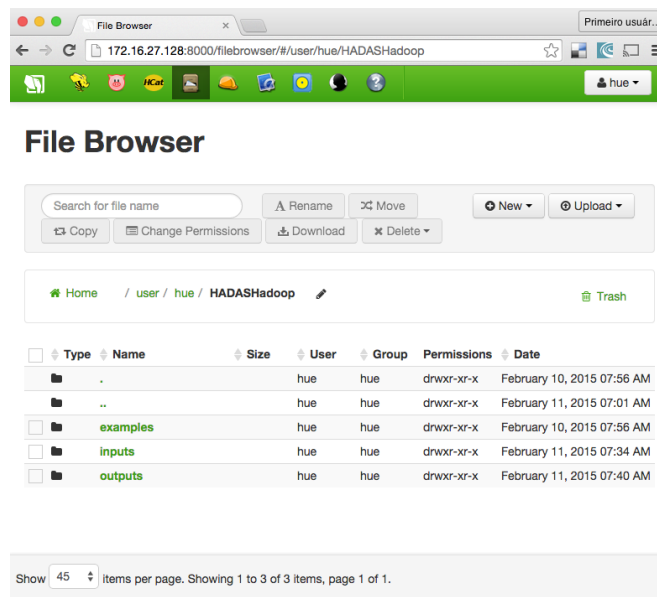


The screenshot shows the HDFS File Browser interface. The browser address bar displays `172.16.27.128:8000/filebrowser/`. The interface includes a search bar, action buttons (Rename, Move, Copy, Change Permissions, Download), and a 'New' button. The breadcrumb path is `Home / user / hue`. A table lists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

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`. The table lists the following items:

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/WeatherTemperature.zip root@172.16.27.128:/usr/lib/
.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:
.DS_Store          filtering-topk.jar  lib/
bin/               inputs/           src/
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 **filtering-topk.jar** 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/filtering-topk.jar /user/hue/HADASHadoop/inputs/designpatterns/Users.xml /user/hue/HADASHadoop/outputs/filtering-topk
```

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



File Browser

Search for file name Rename Move Copy New Upload
Change Permissions Download Delete

[Home](#) / [user](#) / [hue](#) / [HADASHadoop](#) / [outputs](#) Trash

Type	Name	Size	User	Group	Permissions	Date
Folder	.		hue	hue	drwxr-xr-x	February 11, 2015 09:44 AM
Folder	..		hue	hue	drwxr-xr-x	February 10, 2015 07:56 AM
Folder	filtering-topk		root	hue	drwxr-xr-x	February 11, 2015 09:44 AM
Folder	weather		root	hue	drwxr-xr-x	February 11, 2015 09:15 AM



[Home](#) / [user](#) / [hue](#) / [HADASHadoop](#) / [outputs](#) / [filtering-topk](#) / [part-r-00000](#)

ACTIONS
[View As Binary](#)
[Edit File](#)
[Download](#)
[View File](#)
[Location](#)
[Refresh](#)

INFO

Last Modified
Feb. 11, 2015
9:44 a.m.

User
root

Group
hue

Size
11.8 KB

Mode
100644

First Block Previous Block **Next Block** Last Block

Viewing Bytes: - of 12100 (4096 B block size)

```
<row Id="463" Reputation="18249" CreationDate="2011-07-11T08:09:06.287" DisplayName="Michael Dorfman" LastAccessDate="2013-12-23T11:29:04.443" WebsiteUrl="http://www.bokdykk.no" Location="Floro, Norway" AboutMe="What would you like to know?" Views="631" UpVotes="173" DownVotes="88" Age="50" AccountId="4474" />
<row Id="933" Reputation="9185" CreationDate="2011-10-08T17:08:28.203" DisplayName="Mozibur Ullah" LastAccessDate="2014-05-04T02:52:21.773" WebsiteUrl="" Location="London, United Kingdom" AboutMe="I have a Mathematics BA from Oxford, and Theoretical Physics Msc from Imperial College in London. I work as a software engineer, most recently in the city dealing with financial derivatives.<br>I'm interested in Philosophy, Poetry, Theology and Art; and I'm not interested in Gaming, nor SF, nor computers or IT as a way of life.<br>I'm also an occasionally practising artist & poet; my art is probably better than my poetry - although that's getting better. Poets I admire are Yeats, Lorca and Walcott. Do you friend, know what it's all about?<br>The sun, a picture of health hangs still as a citrus fruit on a calm day.<br>Flinty reflections in glass staring back at you,<br>imitating each mellow or awkward move -<br>Narcissus would not have known.<br>Do you friend, know what it's all about?<br>Conversation hiding open-eyed deceit.<br>Mustard oil for hasher-dim fried.<br>Laughter with knives and smiles - for the coup-de-grace.<br>Persophone in spring may know.<br>And will we know what it's all about?<br>Forever losing the living situation, Our inorganic sense drags us under,<br>as our organic sense makes us hunger.<br>Mozibur Ullah, Mar 2013<br>Mozibur Ullah, Mar 2013<br>Mozibur Ullah, Mar 2013" Views="405" UpVotes="924" DownVotes="71" ProfileImageUrl="http://i.stack.imgur.com/sw7YK.jpg" Age="45" AccountId="961274" />
<row Id="233" Reputation="8247" CreationDate="2011-06-16T18:59:35.687" DisplayName="Rex Kerr" LastAccessDate="2014-05-02T02:00:54.527" Views="308" UpVotes="742" DownVotes="69" AccountId="89797" />
```