

Setup for MapReduce Project with Cloudera

1. Download the Cloudera VMWare from
http://www.cloudera.com/content/cloudera/en/downloads/quickstart_vms/cdh-5-4-x.html

The downloaded cloudera VMware has included

1. Cent OS
 2. Hadoop setup (with all libraries)
 3. Java
 4. Eclipse IDE
2. Download the VMware player to start the downloaded cloudera virtual machine
<https://www.vmware.com/products/player>
 3. Open the eclipse which is located in the desktop of the CentOS (The CentOS was loaded by VMware player by using cloudera VMware)
 4. Create a new java project in eclipse (Ranjith'sProject) [**Before going to start this project I was tried WordCount in this development setup**]
 5. Add all Hadoop related dependency libraries to eclipse from below location
/usr/lib/hadoop/client
 6. Create the packages from each and every problem
e.g: com.ranjith.pair, com.ranjith.straip, com.ranjith.hybrid
 7. Create the Main, Mapper, Partitioner and Reducer classes from each every problem with naming convention.
e.g: PairMain, PairMapper, PairPartitioner, PairReducer
 8. Implement all the classes with necessary extends and implements.
 9. Create an input file with given input (e.g: projectInput.txt)
 10. Create HDFS input file by Linux command in terminal by using our input file with same or different name (here I am using the same)

> **hadoop fs -put <inputfile path> <HDFS file path>**

Run the MapReduce program by using Linux command in Terminal

1. In eclipse right click on our project and export the jar file with a specific name (project.jar)
2. Check the HDFS input file by using Linux command which was created in the step 10
>> `hadoop fs -ls`
3. Execute the jar file with created HDFS input by using Linux command in Terminal
 - PairMain class execution:
>> `hadoop jar project.jar com/ranjith/pair/PairMain projectInput.txt`
pairoutput
 - StraipMain class execution:
>> `hadoop jar project.jar com/ranjith/straip/StraipMain projectInput.txt`
straipoutput
 - HybridMain class execution
>> `hadoop jar project.jar com/ranjith/hybrid/HybridMain projectInput.txt`
hybridoutput
4. We can view the output as:
>> `hadoop fs -cat pairoutput`
5. we can view the job status and files and outputs in UI
`http://localhost:50070`
`http://localhost:50070`

cloudera-quickstart-vm-5.4.2-0-vmware - VMware Workstation 12 Player (Non-commercial use only)

Player ▾ | Applications Places System

Hue - Job Browser - Mozilla Firefox

quickstart.cloudera:8888/jobbrowser/

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Job Browser

Logs	ID	Name	Status	User	Maps	Reduces	Queue	Priority	Duration	Submitted
	1442337223975_0013	ranjith_project.jar	SUCCEEDED	cloudera	100%	100%	root.cloudera	N/A	45s	09/17/15 12:40:37
	1442337223975_0012	ranjith_project.jar	SUCCEEDED	cloudera	100%	100%	root.cloudera	N/A	31s	09/17/15 12:36:10
	1442337223975_0011	ranjith_project.jar	SUCCEEDED	cloudera	100%	100%	root.cloudera	N/A	32s	09/17/15 12:34:08
	1442337223975_0010	ranjith_project.jar	FAILED	cloudera	100%	100%	root.cloudera	N/A	1m:38s	09/17/15 12:28:45
	1442337223975_0009	pair.jar	SUCCEEDED	cloudera	100%	100%	root.cloudera	N/A	44s	09/17/15 11:50:36
	1442337223975_0008	pair.jar	SUCCEEDED	cloudera	100%	100%	root.cloudera	N/A	33s	09/17/15 10:07:07
	1442337223975_0007	pair.jar	SUCCEEDED	cloudera	100%	100%	root.cloudera	N/A	1m:15s	09/17/15 10:03:20
	1442337223975_0006	pair.jar	SUCCEEDED	cloudera	100%	100%	root.cloudera	N/A	37s	09/16/15 14:11:56

Hue - Job Browser - M... [cloudera@quickstart:...] Java - Ranjith's-Project...

cloudera-quickstart-vm-5.4.2-0-vmware - VMware Workstation 12 Player (Non-commercial use only)

Player ▾ | Applications Places System

Hue - File Browser - Mozilla Firefox

quickstart.cloudera:8888/filebrowser/

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

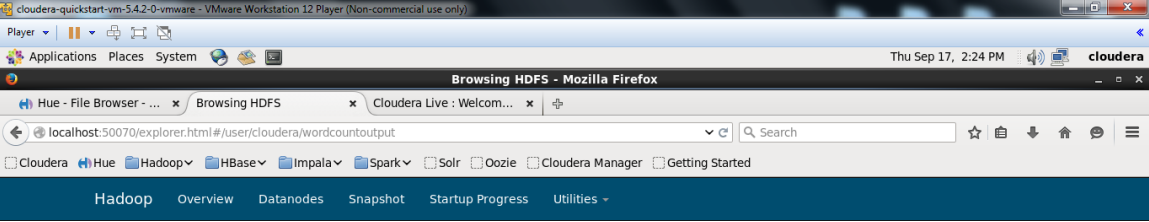
File Browser

Search for file name Actions Move to trash Upload New

Name	Size	User	Group	Permissions	Date
hdfs		hdfs	supergroup	drwxr-xr-x	September 14, 2015 09:55 AM
.		cloudera	cloudera	drwxr-xr-x	September 17, 2015 12:40 PM
.Trash		cloudera	cloudera	drwxr-xr-x	September 16, 2015 02:09 PM
hybridoutput		cloudera	cloudera	drwxr-xr-x	September 17, 2015 12:41 PM
pairoutput		cloudera	cloudera	drwxr-xr-x	September 17, 2015 12:34 PM
projectinput.txt	60 bytes	cloudera	cloudera	-rw-r--r--	September 17, 2015 11:48 AM
straioutput		cloudera	cloudera	drwxr-xr-x	September 17, 2015 12:36 PM
wordcountoutput		cloudera	cloudera	drwxr-xr-x	September 15, 2015 10:02 AM

Show 45 of 6 items Page 1 of 1

Hue - File Browser - M... [cloudera@quickstart:...] Java - Ranjith's-Project...



Browse Directory

Permission	Owner	Group	Size	Replication	Block Size	Name
-rw-r--r--	cloudera	cloudera	0 B	1	128 MB	_SUCCESS
-rw-r--r--	cloudera	cloudera	30 B	1	128 MB	part-00000

Hadoop, 2014.

