

# CS-5200

## Homework #2: WordCount

Name of Student : Anand

A-Number : A02091430

-----

**Output :**

Number	File Name	Total number of words	Performance
1	Les Miserables.txt	68117	1 min, 25 sec
2	Clarissa Harlowe; or the history of a young lady.txt	110042	46 sec
3	The Complete Works of William Shakespeare.txt	124788	1 min, 17 sec
4	decline and fall of the roman empire.txt	130693	1 min, 00 sec

“TOTAL\_NUMBER\_OF\_WORDS” is the first field in each output file, which gives total number of words, followed by the word occurrences.

To get the Performance time, I used hadoop UI.

<http://localhost:8088/cluster>

In the details of the job performed, we can get the time taken by the application and aggregate resource allocation. It has shown in the following snapshots.

## 1. Les Miserables.txt

The screenshot shows the Cloudera Manager interface in a Mozilla Firefox browser. The address bar displays `localhost:8088/cluster/app/application_1428104399335_0003`. The left sidebar contains a 'Cluster' menu with options like 'About', 'Nodes', 'Applications', and 'Scheduler'. The main content area is titled 'Application Overview' and displays the following details:

- User:** cloudera
- Name:** wordcount
- Application Type:** MAPREDUCE
- Application Tags:**
- State:** FINISHED
- FinalStatus:** SUCCEEDED
- Started:** Fri Apr 03 18:06:12 -0700 2015
- Elapsed:** 1mins, 25sec
- Tracking URL:** [History](#)
- Diagnostics:**

Below the overview, the 'Application Metrics' section shows:

- 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:** 324256 MB-seconds, 223 vcore-seconds

The bottom of the screen shows the taskbar with various application icons and a system clock indicating 6:22 PM on 4/3/2015.

## 2. Clarissa Harlowe; or the history of a young lady.txt

The screenshot shows the Cloudera Manager interface in a Mozilla Firefox browser. The address bar displays `localhost:8088/cluster/app/application_1428104399335_0004`. The left sidebar contains a 'Cluster' menu with options like 'About', 'Nodes', 'Applications', and 'Scheduler'. The main content area is titled 'Application Overview' and displays the following details:

- User:** cloudera
- Name:** wordcount
- Application Type:** MAPREDUCE
- Application Tags:**
- State:** FINISHED
- FinalStatus:** SUCCEEDED
- Started:** Fri Apr 03 18:08:59 -0700 2015
- Elapsed:** 46sec
- Tracking URL:** [History](#)
- Diagnostics:**

Below the overview, the 'Application Metrics' section shows:

- 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:** 172635 MB-seconds, 114 vcore-seconds

At the bottom, the 'ApplicationMaster' section displays a table with the following columns: Attempt Number, Start Time, Node, and Logs. The first row shows Attempt 1 starting at Fri Apr 03 18:08:59 -0700 2015 on node quickstart-cloudera-8943.

### 3. The Complete Works of William Shakespeare.txt

The screenshot shows the Cloudera Manager interface in a Mozilla Firefox browser. The main panel displays the details of a MapReduce application named 'wordcount'. The application has finished successfully. The left sidebar shows the 'Applications' menu with various status filters. The bottom of the interface shows the 'ApplicationMaster' table with one attempt.

**Application Details:**

- User: cloudera
- Name: wordcount
- Application Type: MAPREDUCE
- Application Tags:
- State: FINISHED
- FinalStatus: SUCCEEDED
- Started: Fri Apr 03 18:11:13 -0700 2015
- Elapsed: 1mins, 17sec
- Tracking URL: History
- 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: 291536 MB-seconds, 200 vcore-seconds

**ApplicationMaster Table:**

Attempt Number	Start Time	Node	Logs
1	Fri Apr 03 18:11:13 -0700 2015	quickstart cloudera-8943	log

### 4. decline and fall of the roman empire.txt

The screenshot shows the Cloudera Manager interface in a Mozilla Firefox browser. The main panel displays the details of a MapReduce application named 'wordcount'. The application has finished successfully. The left sidebar shows the 'Applications' menu with various status filters. The bottom of the interface shows the 'ApplicationMaster' table with one attempt.

**Application Details:**

- User: cloudera
- Name: wordcount
- Application Type: MAPREDUCE
- Application Tags:
- State: FINISHED
- FinalStatus: SUCCEEDED
- Started: Fri Apr 03 18:13:48 -0700 2015
- Elapsed: 1mins, 0sec
- Tracking URL: History
- 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: 227562 MB-seconds, 153 vcore-seconds

**ApplicationMaster Table:**

Attempt Number	Start Time	Node	Logs
1	Fri Apr 03 18:13:48 -0700 2015	quickstart cloudera-8943	log