**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](https://usu.instructure.com/courses/351921/files/56333783/download?wrap=1) | 68117 | 1 min, 25 sec |
| 2 | [Clarissa Harlowe; or the history of a young lady.txt](https://usu.instructure.com/courses/351921/files/56333792/download?wrap=1) | 110042 | 46 sec |
| 3 | [The Complete Works of William Shakespeare.txt](https://usu.instructure.com/courses/351921/files/56333796/download?wrap=1) | 124788 | 1 min, 17 sec |
| 4 | [decline and fall of the roman empire.txt](https://usu.instructure.com/courses/351921/files/56333806/download?wrap=1) | 130693 | 1 min, 00 sec |

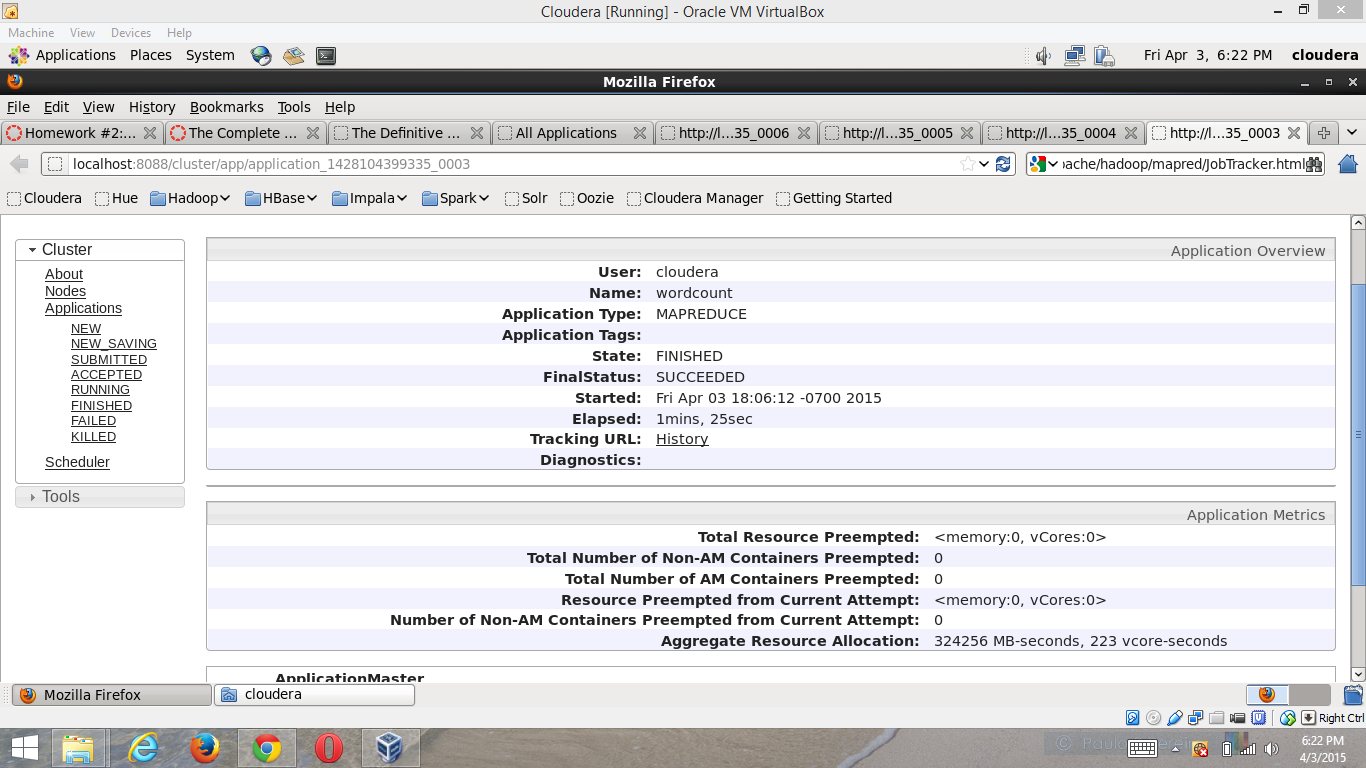
“**TOTAL\_NUMBER\_OF\_WORDS**” is the first field in each output file, which gives total number of words, followed by the word occurances.

To get the Performance time, I used hadoop UI.

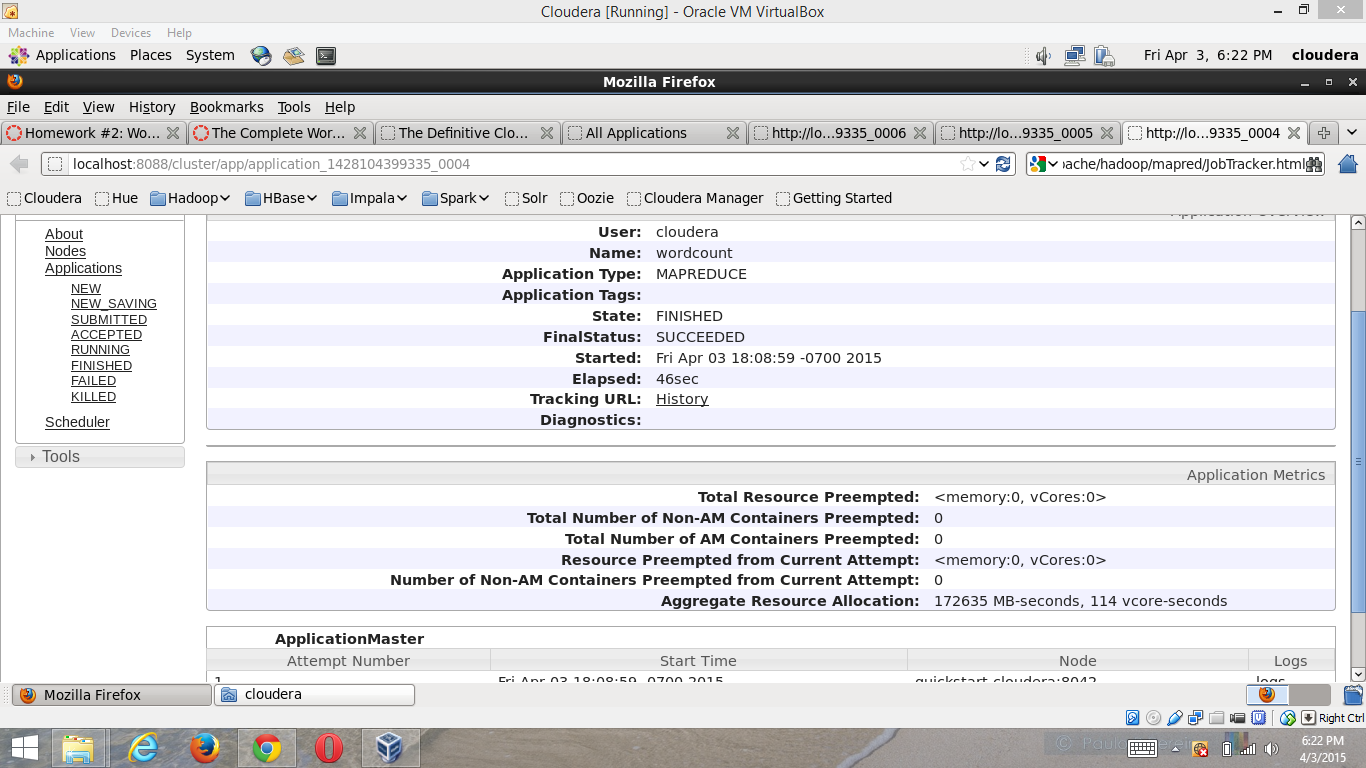
<http://localhost:8088/cluster>

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

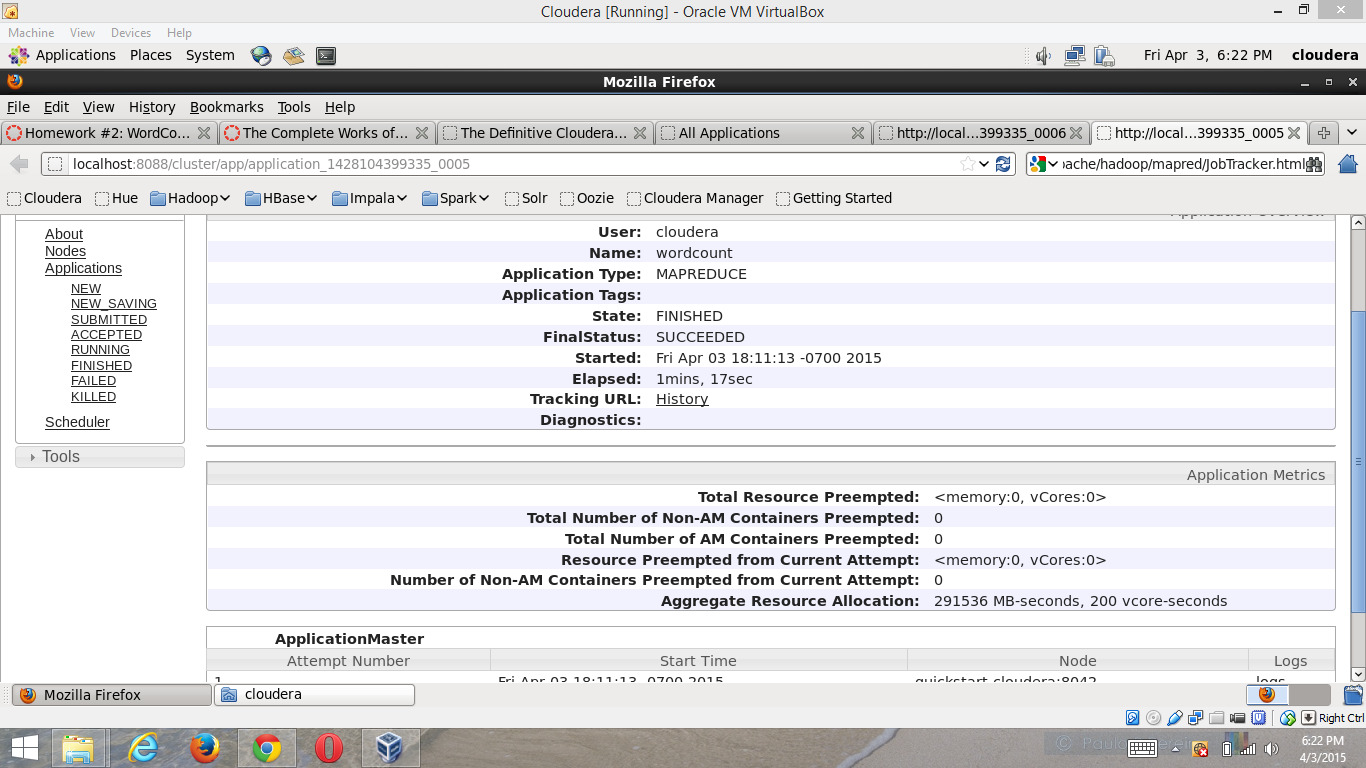
1. [**Les Miserables.txt**](https://usu.instructure.com/courses/351921/files/56333783/download?wrap=1)

****

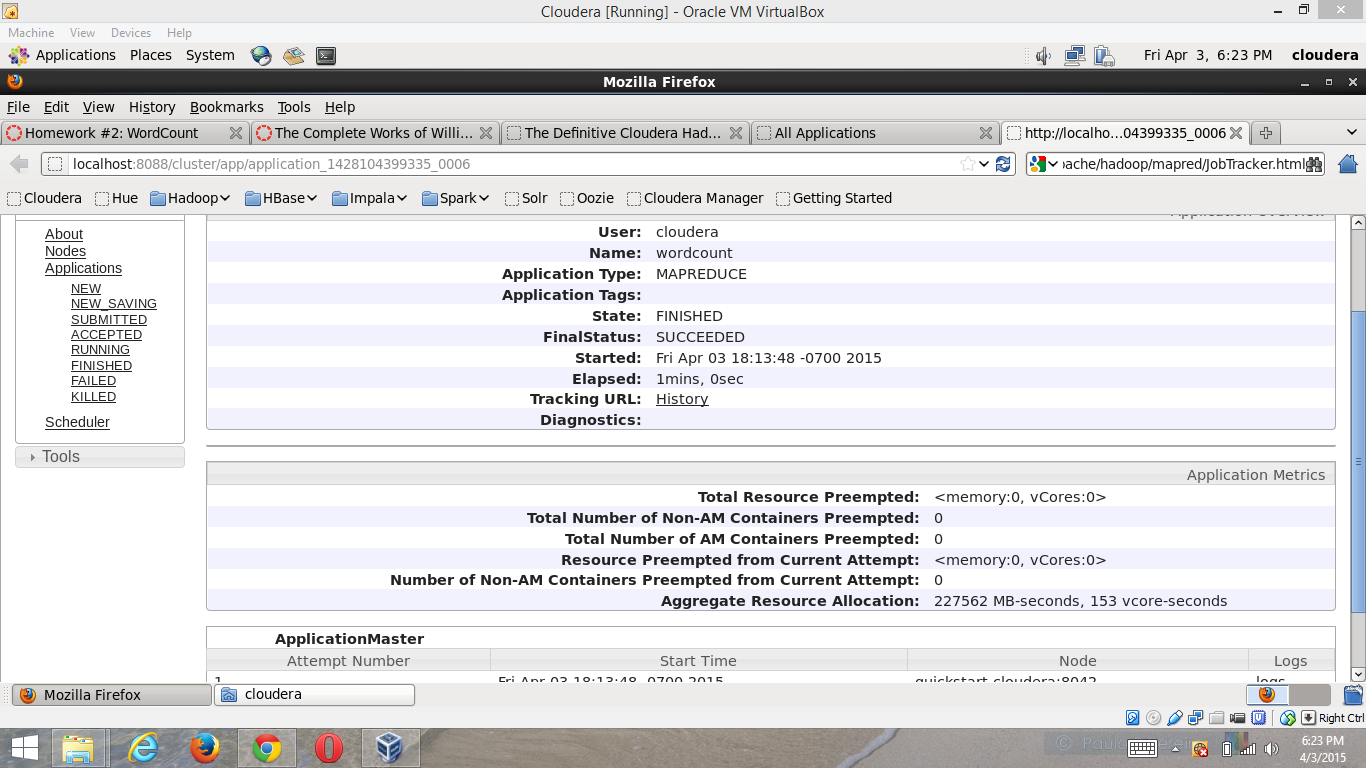
1. [**Clarissa Harlowe; or the history of a young lady.txt**](https://usu.instructure.com/courses/351921/files/56333792/download?wrap=1)

****

1. [**The Complete Works of William Shakespeare.txt**](https://usu.instructure.com/courses/351921/files/56333796/download?wrap=1)

****

1. [**decline and fall of the roman empire.txt**](https://usu.instructure.com/courses/351921/files/56333806/download?wrap=1)

****