

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG



BÁO CÁO MÔN HỌC
Lưu trữ và xử lý dữ liệu lớn

LAB02

Nhóm: **RHUST**

Phan Thành Đạt	20173001
Hoàng Văn Chương	20172984
Đỗ Minh Vũ	20173471
Nguyễn Thị Bắc	20172963

Number Mapper = (Total data size)/ (Input split size)

Trong đó: Input split size = max (minimumSize, min (maximumSize, blockSize))

Giá trị mặc định: minimumSize=0

maximumSize=Long.MaxValue

1) TH1: Split input size hdfs = HDFS block size (default)

3GB data, block size = 128 MB => Number Mapper = 24




Google Chrome Thg 10 31 22:49

Apache Hadoop x Browsing HDFS x How to Install x All Applications x About the Team x (1) Facebook x calculator - TI x







Not secure | node-master:9870/explorer.html#/books

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Browse Directory

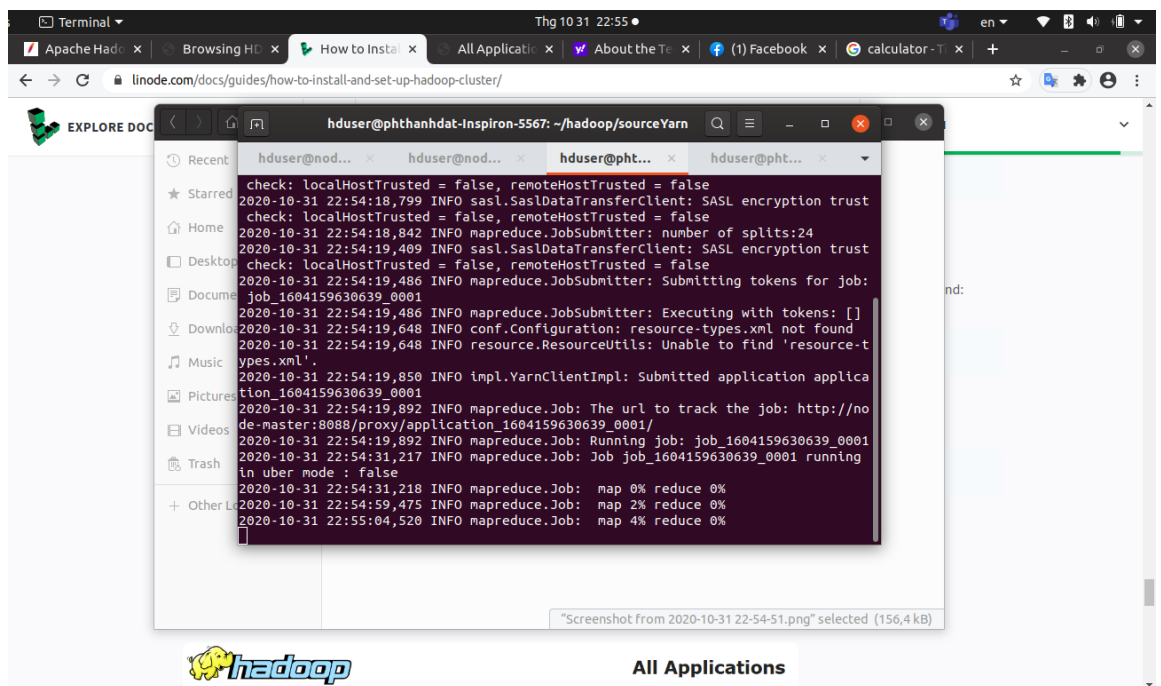
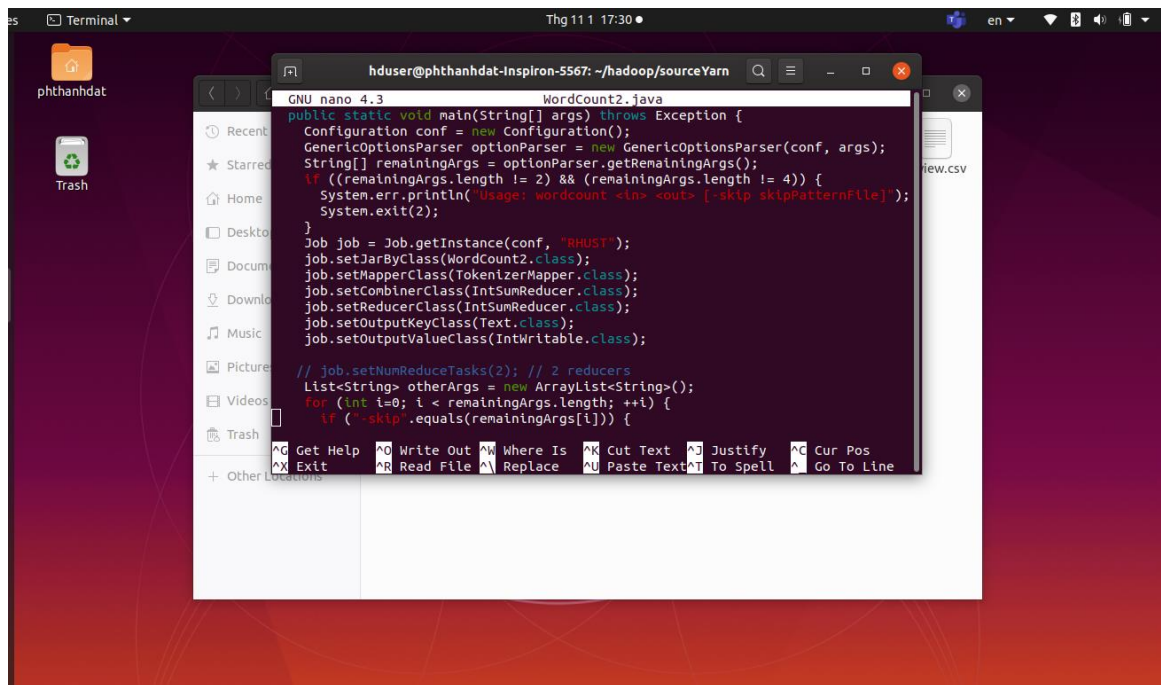
/books Go!   

Show 25 entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	hduser	supergroup	170.6 KB	Oct 27 10:26	2	1 MB	alice.txt	
<input type="checkbox"/>	-rw-r--r--	hduser	supergroup	3 GB	Oct 31 22:43	2	128 MB	bigfile.txt	
<input type="checkbox"/>	-rw-r--r--	hduser	supergroup	440.22 KB	Oct 27 10:26	2	1 MB	frankenstein.txt	
<input type="checkbox"/>	-rw-r--r--	hduser	supergroup	593.55 KB	Oct 27 10:26	2	1 MB	holmes.txt	
<input type="checkbox"/>	-rw-r--r--	hduser	supergroup	2.07 MB	Oct 27 10:29	2	1 MB	sample-2mb-text-file.txt	

Showing 1 to 5 of 5 entries Previous 1 Next

Hadoop, 2019.



Firefox Web Browser Thg 10 31 22:54

MapReduce Application application_1604159630639_0001 - Mozilla Firefox

MapReduce Application app: X

node-master:8088/proxy/application_1604159630639_0001/

hadoop

MapReduce Application
application_1604159630639_0001

Cluster

Application

About
Jobs

Tools

Active Jobs

Show 20 entries

Search:

Job ID	Name	State	Map Progress	Maps Total	Maps Completed	Reduce Progress	Reduces Total	Reduces Completed
job_1604159630639_0001	RHUST	RUNNING		24	0		1	0

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

Logged in as: dr.who

Firefox Web Browser Thg 10 31 23:00

MapReduce Application application_1604159630639_0001 - Mozilla Firefox

MapReduce Application app: X

node-master:8088/proxy/application_1604159630639_0001/

hadoop

MapReduce Application
application_1604159630639_0001

Cluster

Application

About
Jobs

Tools

Active Jobs

Show 20 entries

Search:

Job ID	Name	State	Map Progress	Maps Total	Maps Completed	Reduce Progress	Reduces Total	Reduces Completed
job_1604159630639_0001	RHUST	RUNNING		24	17		1	0

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

Logged in as: dr.who

Google Chrome Thg 10 31 23:01

Apache Hadoop x Browsing HD x How to Insta x All Applicatio x About the Te x Facebook x calculator - TI x

Not secure | node-master:8088/cluster

hadoop

All Applications

Cluster

- About
- Nodes
- Node Labels
- Applications
- NEW
- NEW_SAVING
- SUBMITTED
- ACCEPTED
- RUNNING
- FINISHED
- FAILED
- KILLED
- Scheduler
- Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Mem
1	0	1	0	2	2.50 GB	3 GB

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes
2	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Rur Cont
application_1604159630639_0001	hduser	RHUST	MAPREDUCE	default	0	Sat Oct 31 22:54:19 +0700 2020	Sat Oct 31 22:54:20 +0700 2020	N/A	RUNNING	UNDEFINED	2

Showing 1 to 1 of 1 entries

Terminal Thg 10 31 23:08

Apache Hadoop x Browsing HD x How to Insta x All Applicatio x About the Te x (1) Facebook x calculator - TI x

Not secure | node-master:9870/explorer.html#/books

Hadoop Overview

Browse Dir

/books

Show 25 entries

Permission

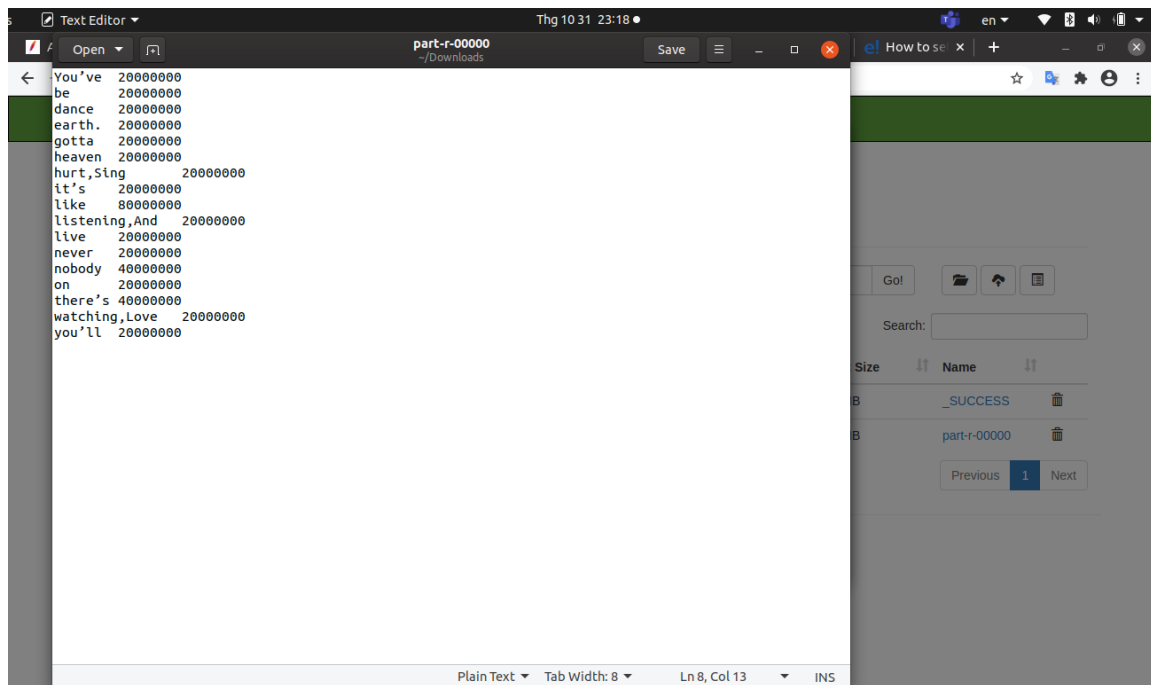
Showing 1 to 5 of 5 entries

Hadoop, 2019.

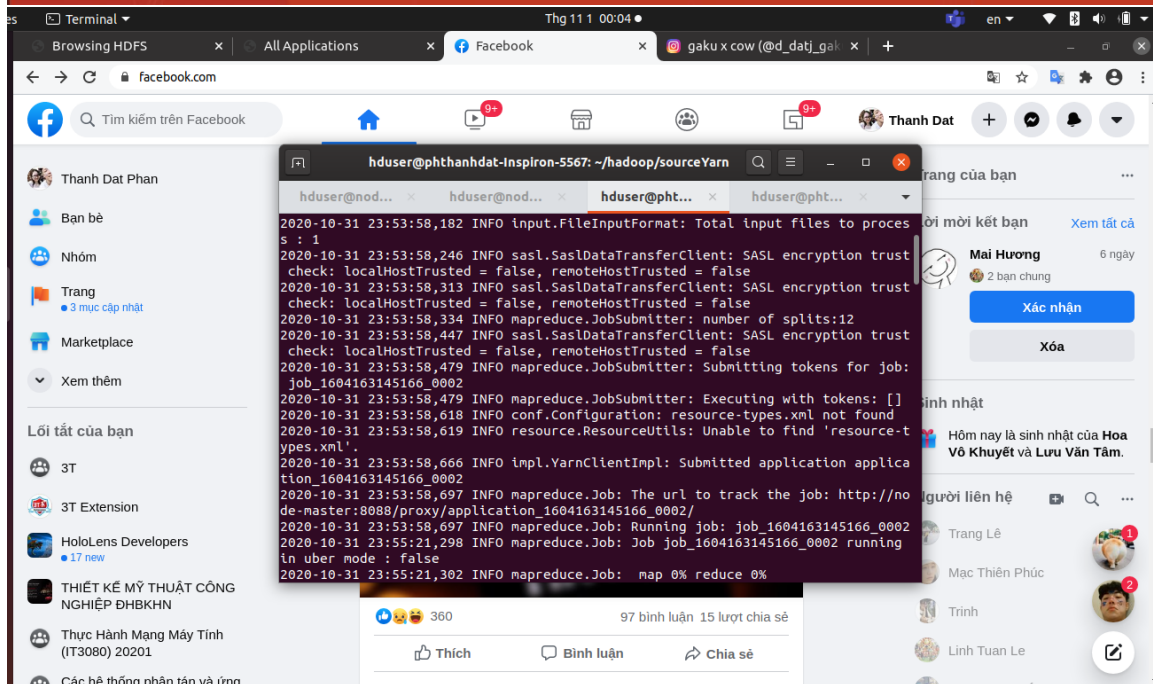
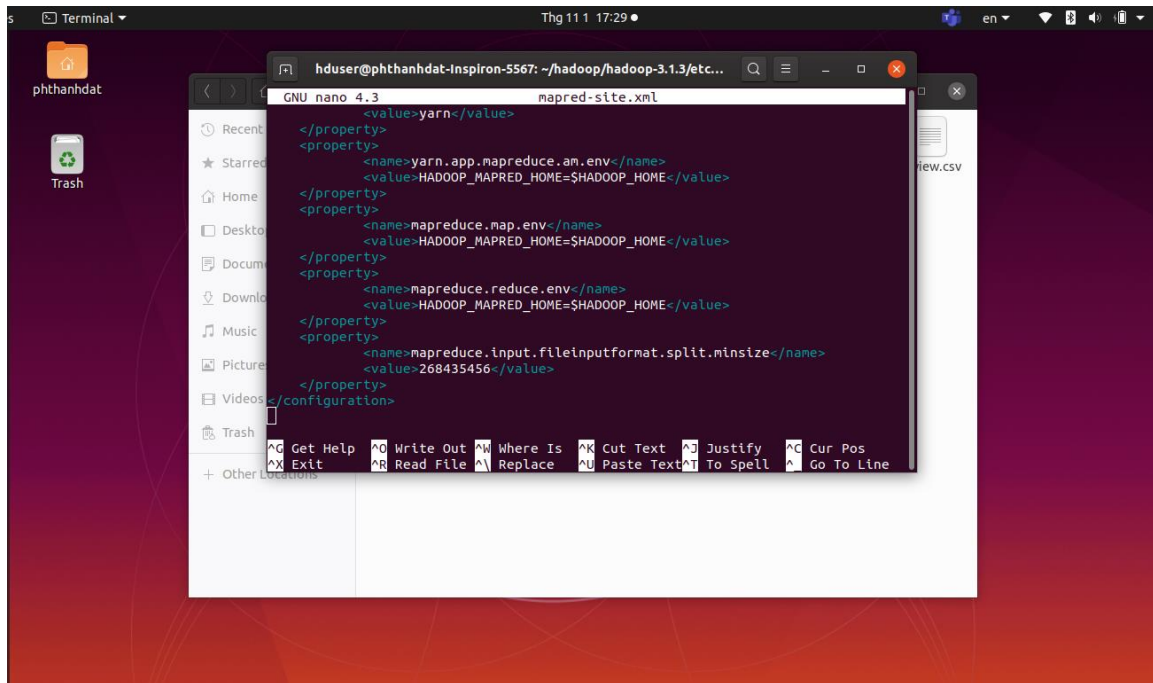
```

hduser@phthanhdai-Inspiron-5567: ~/hadoop/sourceYarn
hduser@nod... x hduser@nod... x hduser@pht... x hduser@pht... x
HDFS: Number of write operations=2
Job Counters
Killed map tasks=1
Launched map tasks=24
Launched reduce tasks=1
Data-local map tasks=24
Total time spent by all maps in occupied slots (ms)=453986
Total time spent by all reduces in occupied slots (ms)=4095
Total time spent by all map tasks (ms)=453986
Total time spent by all reduce tasks (ms)=4095
Total vcore-milliseconds taken by all map tasks=453986
Total vcore-milliseconds taken by all reduce tasks=4095
Total megabyte-milliseconds taken by all map tasks=464881664
Total megabyte-milliseconds taken by all reduce tasks=4193280
Map-Reduce Framework
Map input records=20000000
Map output records=440000000
Map output bytes=4960000000
Map output materialized bytes=5664
Input split bytes=2544
Combine input records=440002448
Combine output records=2856
Reduce input groups=17

```



2) TH2: Thay đổi số lượng mapper bằng thuộc tính minsize input




Firefox Web Browser Thg 10 31 23:55

MapReduce Application application_1604163145166_0002 - Mozilla Firefox

MapReduce Application app X +

node-master:8088/proxy/application_1604163145166_0002/

Logged in as: dr.who



MapReduce Application application_1604163145166_0002

Cluster

Application

About Jobs

Tools

Active Jobs

Show 20 entries

Job ID	Name	State	Map Progress	Maps Total	Maps Completed	Reduce Progress	Reduces Total	Reduces Completed
job_1604163145166_0002	RHUST	RUNNING		12	0		1	0

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

Terminal Thg 11 1 00:04

Browsing HDFS All Applications Facebook gaku x cow (@_datj_gak

facebook.com

Thanh Dat Phan

Ban bè

Nhóm

Trang 3 mục cập nhật

Marketplace

Xem thêm

Lỗi tắt của bạn

3T

3T Extension

HoloLens Develo 17 new

THIẾT KẾ MỸ THUẬT NGHIỆP ĐHBKHN

Thực Hành Mạng Máy Tính (IT3080) 20201

Các hệ thống phân tán và ứng

Recent

Starred

Home

Desktop

Documents

Downloads

Music

Pictures

Videos

Trash

Other Location

```
hduser@phthanhdatt-nsprion-5567: ~/hadoop/sourceYarn
hduser@nod... x hduser@nod... x hduser@pht... x hduser@pht... x
GC time elapsed (ms)=6382
CPU time spent (ms)=365180
Physical memory (bytes) snapshot=2871177216
Virtual memory (bytes) snapshot=32488681472
Total committed heap usage (bytes)=2077081600
Peak Map Physical memory (bytes)=231866368
Peak Map Virtual memory (bytes)=2498691072
Peak Reduce Physical memory (bytes)=119308288
Peak Reduce Virtual memory (bytes)=2504388608

Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
WordCount2$TokenizerMapper$CountersEnum
INPUT_WORDS=440000000
File Input Format Counters
Bytes Read=3220045056
File Output Format Counters
Bytes Written=281
hduser@phthanhdatt-nsprion-5567:~/hadoop/sourceYarn$
```

"Screenshot from 2020-10-31 23-55-35.png" selected (156,4 kB)

Thích Bình luận Chia sẻ

3) Thêm một số tùy chọn

The screenshot shows a web browser window with a sidebar on the left and a main content area. The sidebar contains a list of Latin words and their corresponding frequencies. The main content area is mostly blank, with a search bar at the top and a list of items below it.

Word	Frequency
A	372
Ac	431
Accumsan	176
Adipiscing	363
Aenean	177
Aliquam	436
Aliquet	313
Amet	816
Ante	76
Arcu	453
At	624
Auctor	191
Augue	241
Bibendum	223
Blandit	213
Commodo	280
Condimentum	149
Congue	148
Consectetur	235
Consequat	227
Convallis	200
Cras	245
Cum	38
Curabitur	77
Cursus	314
Dapibus	36
Diam	514
Dictum	176
Dictumst	69
Dignissim	238
Dis	29
Dolor	248
Donec	250
Dui	262
Duis	231
Egestas	524
Eget	690

The main content area shows a search bar with the text "Search:" and a list of items below it. The items are:

- _SUCCESS
- part-r-00000

At the bottom of the main content area, there are buttons for "Previous", "1", and "Next".

Thg 10/31 22:15

Apache Hadoop x Browsing HDFS x How to Install an x All Applications x About the Test D x Facebook x + - o x

hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html

Sample Runs

Sample text-files as input:

```
$ bin/hadoop
/user/joe/w
/user/joe/w

$ bin/hadoop
Hello World

$ bin/hadoop
Hello Hadoop
```

Run the application:

```
$ bin/hadoop
```

Output:

```
$ bin/hadoop
Bye 1
Goodbye 1
Hadoop, 1
Hello 2
World! 1
World, 1
hadoop.. 1
to 1
```

Notice that the input

Now, lets plug-in a pattern-file which lists the word-patterns to be ignored, via the DistributedCache.

```
$ bin/hadoop fs -cat /user/joe/wordcount/patterns.txt
\
\
\
```

Google Chrome Thg 10 31 22:17

Apache Hadoop x Browsing HDFS x How to Install an x All Applications x About the Test D x Facebook x

Not secure | node-master:8088/cluster

Cluster metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total
5	0	5	0	0 B	3 GB	

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes
2	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Max
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:1536, vCores:1>

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers
application_1604139774142_0006	hduser	RHUST	MAPREDUCE	default	0	Sat Oct 31 22:15:44 +0700 2020	Sat Oct 31 22:15:44 +0700 2020	Sat Oct 31 22:16:08 +0700 2020	FINISHED	SUCCEEDED	N/A
application_1604139774142_0005	hduser	RHUST	MAPREDUCE	default	0	Sat Oct 31 17:38:25 +0700 2020	Sat Oct 31 17:38:25 +0700 2020	Sat Oct 31 17:38:48 +0700 2020	FINISHED	SUCCEEDED	N/A
application_1604139774142_0003	hduser	word count	MAPREDUCE	default	0	Sat Oct 31 17:35:10 +0700 2020	Sat Oct 31 17:35:11 +0700 2020	Sat Oct 31 17:35:33 +0700 2020	FINISHED	SUCCEEDED	N/A
application_1604139774142_0002	hduser	word count	MAPREDUCE	default	0	Sat Oct 31 17:27:06 +0700 2020	Sat Oct 31 17:27:06 +0700 2020	Sat Oct 31 17:28:40 +0700 2020	FINISHED	FAILED	N/A
application_1604139774142_0001	hduser	word count	MAPREDUCE	default	0	Sat Oct 31 17:24:34 +0700 2020	Sat Oct 31 17:24:34 +0700 2020	Sat Oct 31 17:26:12 +0700 2020	FINISHED	FAILED	N/A


Firefox Web Browser Thg 10 31 22:16

MapReduce Application application_1604139774142_0006 - Mozilla Firefox

MapReduce Application app: X

node-master:8088/proxy/application_1604139774142_0006/

Logged in as: dr.who

 **MapReduce Application**
application_1604139774142_0006

Cluster

Application

About Jobs

Tools

Active Jobs

Show 20 entries

Job ID	Name	State	Map Progress	Maps Total	Maps Completed	Reduce Progress	Reduces Total	Reduces Completed
job_1604139774142_0006	RHUST	RUNNING		2	2		1	0

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

