Information Technology Institute

Mapreduce Code

Prepared by Dr. Ahmed Sallam

Slides based on original course by UDACITY

INPUT DATA

INPUT DATA

2012-01-01 12:01 San Jose Music 12.99 Amex

INPUT DATA

2012-01-01 — 12:01 — San Jose — Music — 12.99 — Amex

How To Find Total Sales / Store?

How to find total sales / store?

KEY	VALUE
time	store name

How to find total sales / store?

	KEY	VALUE
☐	time	store name
☐	cost	store name

How to find total sales / store?

	KEY	VALUE
☐	time	store name
☐	cost	store name
☐	store name	cost

How to find total sales / store?

	KEY	VALUE
☐	time	store name
☐	cost	store name
☐	store name	cost
☐	store name	product type

How To FIND TOTAL SALES / STORE?

2012-01-01 12:01 San Jose Music 12.99 Amex

	KEY	VALUE
☐	time	store name
☐	cost	store name
☒	store name	cost
☐	store name	product type

MAPPER

```
def mapper():
    for line in sys.stdin:
        data = line.strip().split("\t")
        date, time, store, item, cost, payment = data
        print "{0}\t{1}".format(store, cost)
```

MAPPER

```
def mapper():
    for line in sys.stdin:
        data = line.strip().split("\t")
        date, time, store, item, cost, payment = data
        print "{0}\t{1}".format(store, cost)
```

```
2012-01-01  12:01   San Jose    Music   12.99   Amex
2012-01-02  There was an error trying to connect to the database. Please try again.
```

```
def mapper():

    for line in sys.stdin:

        data = line.strip().split("\t")

        date, time, store, item, cost, payment = data

        print "{0}\t{1}".format(store, cost)
```

```
def mapper():
    for line in sys.stdin:
        data = line.strip().split("\t")

        if len(data) == 6:

            date, time, store, item, cost, payment = data

            print "{0}\t{1}".format(store, cost)
```

WHAT HAPPENS BETWEEN MAPPER AND REDUCER?

WHAT HAPPENS BETWEEN MAPPER AND REDUCER?

- BUBBLE SORT

WHAT HAPPENS BETWEEN MAPPER AND REDUCER?

- ☐ BUBBLE SORT
- ☐ SHUFFLE AND SORT

WHAT HAPPENS BETWEEN MAPPER AND REDUCER?

- ☐ BUBBLE SORT
- ☐ SHUFFLE AND SORT
- ☐ FIND AND SORT

WHAT HAPPENS BETWEEN MAPPER AND REDUCER?

- ☐ BUBBLE SORT
- ☐ SHUFFLE AND SORT
- ☐ FIND AND SORT
- ☐ QUICKSORT

REDUCING

REDUCING

Miami	12.34
Miami	99.07
Miami	3.14
NYC	99.77
NYC	88.99

REDUCING

```
Miami   12.34
Miami   99.07
Miami   3.14
NYC     99.77
NYC     88.99
```

HADOOP STREAMING

REDUCING

```
Miami   12.34
Miami   99.07
Miami   3.14
NYC     99.77
NYC     88.99
```

HADOOP STREAMING

↳ ANY LANGUAGE

KEEPING TRACK

Miami	12.34
Miami	99.07
Miami	3.14
NYC	99.77
NYC	88.99

KEEPING TRACK

Miami	12.34
Miami	99.07
Miami	3.14
NYC	99.77
NYC	88.99

□ PREVIOUS COST

KEEPING TRACK

Miami	12.34
Miami	99.07
Miami	3.14
NYC	99.77
NYC	88.99

- [ ] PREVIOUS COST
- [ ] CURRENT COST

KEEPING TRACK

Miami	12.34
Miami	99.07
Miami	3.14
NYC	99.77
NYC	88.99

- ☐ PREVIOUS COST
- ☐ CURRENT COST
- ☐ TOTAL SALES PER STORE

KEEPING TRACK

Miami	12.34
Miami	99.07
Miami	3.14
NYC	99.77
NYC	88.99

- [ ] PREVIOUS COST
- [ ] CURRENT COST
- [ ] TOTAL SALES PER STORE
- [ ] PREVIOUS STORE

KEEPING TRACK

Miami	12.34
Miami	99.07
Miami	3.14
NYC	99.77
NYC	88.99

- ☐ PREVIOUS COST
- ☐ CURRENT COST
- ☐ TOTAL SALES PER STORE
- ☐ PREVIOUS STORE
- ☐ CURRENT STORE

KEEPING TRACK

Miami	12.34
Miami	99.07
Miami	3.14
NYC	99.77
NYC	88.99

- [ ] PREVIOUS COST
- [x] CURRENT COST
- [ ] TOTAL SALES PER STORE
- [ ] PREVIOUS STORE
- [ ] CURRENT STORE

KEEPING TRACK

Miami	12.34
Miami	99.07
Miami	3.14
NYC	99.77
NYC	88.99

- [ ] PREVIOUS COST
- [x] CURRENT COST
- [x] TOTAL SALES PER STORE
- [ ] PREVIOUS STORE
- [ ] CURRENT STORE

KEEPING TRACK

Miami	12.34
Miami	99.07
Miami	3.14
NYC	99.77
NYC	88.99

- [ ] PREVIOUS COST
- [x] CURRENT COST
- [x] TOTAL SALES PER STORE
- [ ] PREVIOUS STORE
- [x] CURRENT STORE

KEEPING TRACK

Miami	12.34
Miami	99.07
Miami	3.14
NYC	99.77
NYC	88.99

- [ ] PREVIOUS COST
- [x] CURRENT COST
- [x] TOTAL SALES PER STORE
- [x] PREVIOUS STORE
- [x] CURRENT STORE

REDUCER CODE

```
Miami   12.34
Miami   99.07
Miami   3.14
NYC     99.77
NYC     88.99
```

```
def reducer():

    salesTotal = 0
    oldKey = None

    for line in sys.stdin:
        data = line.strip().split("\t")

        if len(data) != 2:
            continue

        thisKey, thisSale = data

        if oldKey and oldKey != thisKey:
            print "{0}\t{1}".format(oldKey, salesTotal)

            salesTotal = 0

        oldKey = thisKey
        salesTotal += float(thisSale)
```

REDUCER CODE

ARE WE DONE?

○ YES

○ NO

```
def reducer():

    salesTotal = 0
    oldKey = None

    for line in sys.stdin:
        data = line.strip().split("\t")

        if len(data) != 2:
            continue

        thisKey, thisSale = data

        if oldKey and oldKey != thisKey:
            print "{0}\t{1}".format(oldKey, salesTotal)

            salesTotal = 0

        oldKey = thisKey
        salesTotal += float(thisSale)
```

REDUCER CODE

ARE WE DONE?

○ YES

⊘ NO

```
def reducer():

    salesTotal = 0
    oldKey = None

    for line in sys.stdin:
        data = line.strip().split("\t")

        if len(data) != 2:
            continue

        thisKey, thisSale = data

        if oldKey and oldKey != thisKey:
            print "{0}\t{1}".format(oldKey, salesTotal)

            salesTotal = 0

        oldKey = thisKey
        salesTotal += float(thisSale)
```

REDUCER CODE

ARE WE DONE?

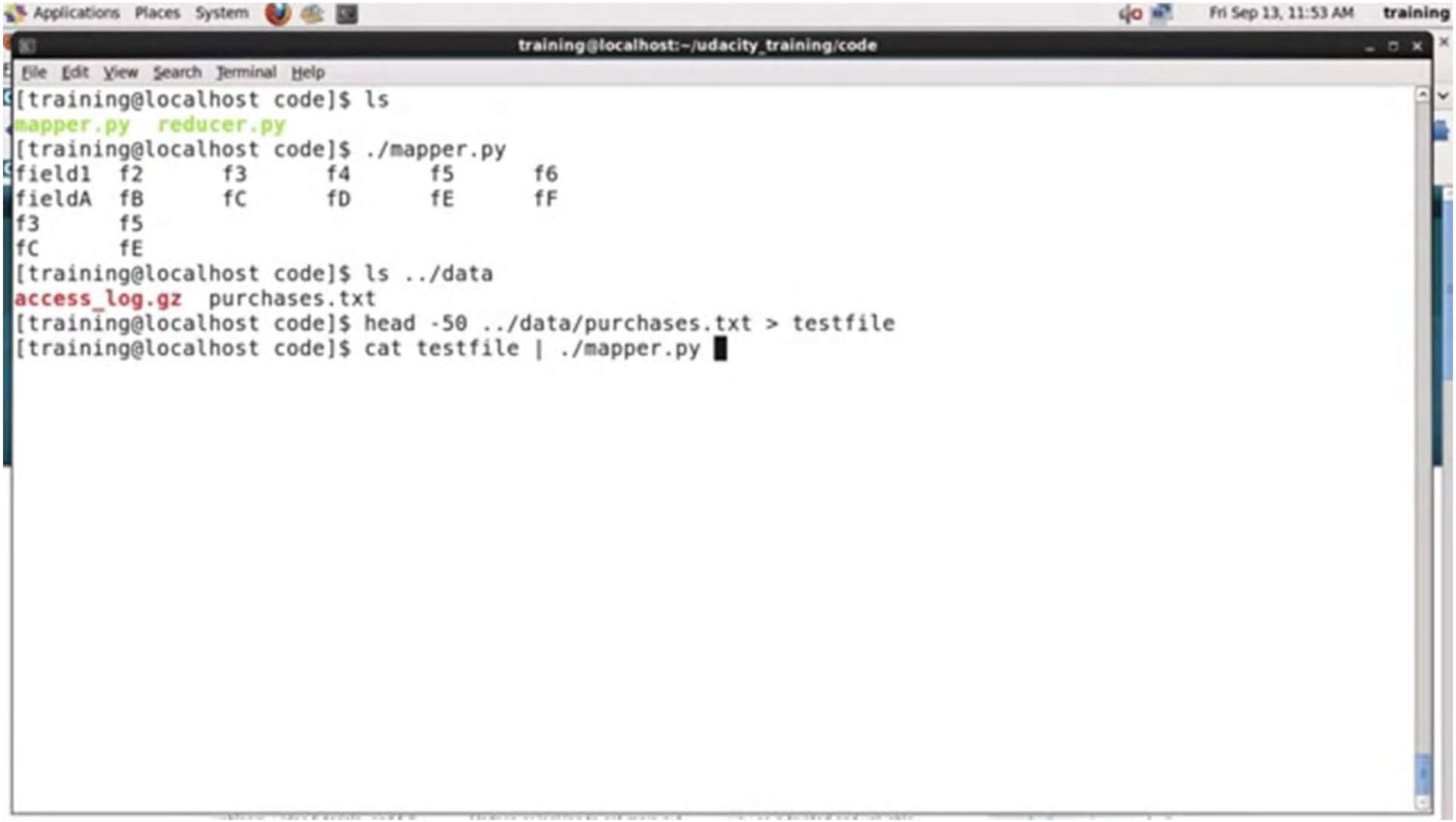
○ YES

⊘ NO

```
def reducer():

    salesTotal = 0
    oldKey = None

    for line in sys.stdin:
        data = line.strip().split("\t")

        if len(data) != 2:
            continue

        thisKey, thisSale = data

        if oldKey and oldKey != thisKey:
            print "{0}\t{1}".format(oldKey, salesTotal)

            salesTotal = 0

        oldKey = thisKey
        salesTotal += float(thisSale)

    if oldKey != None:
        print "{0}\t{1}".format(oldKey, salesTotal)
```


Applications Places System
Fri Sep 13, 11:53 AM training
training@localhost:~/udacity_training/code
File Edit View Search Terminal Help
[training@localhost code]$ ls
mapper.py reducer.py
[training@localhost code]$

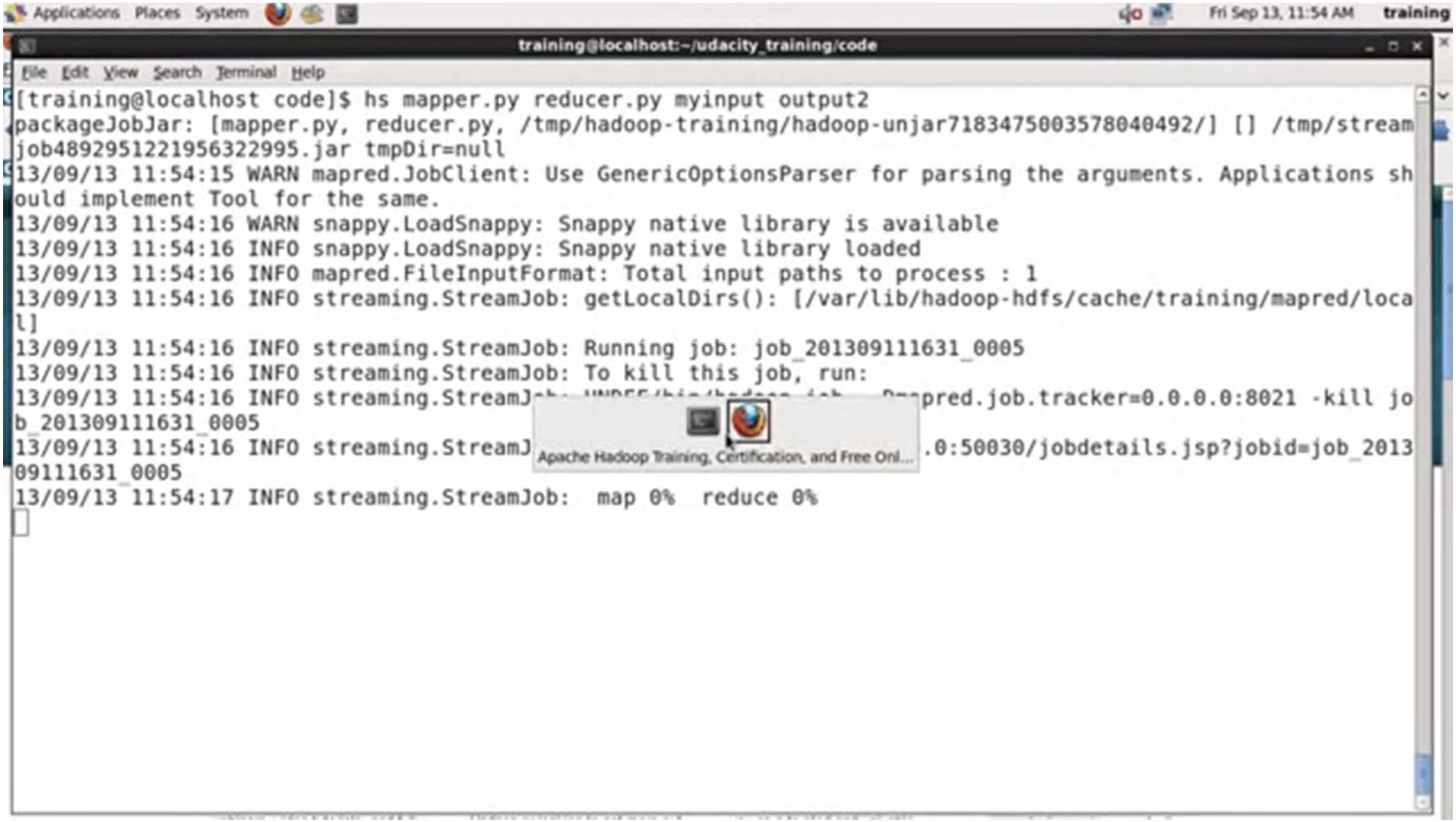

Applications Places System
Fri Sep 13, 11:53 AM training
training@localhost:~/udacity_training/code
File Edit View Search Terminal Help
[training@localhost code]$ ls
mapper.py reducer.py
[training@localhost code]$ ./mapper.py
field1 f2 f3 f4 f5 f6
fieldA fB fC fD fE fF
f3 f5
fC fE
[training@localhost code]$

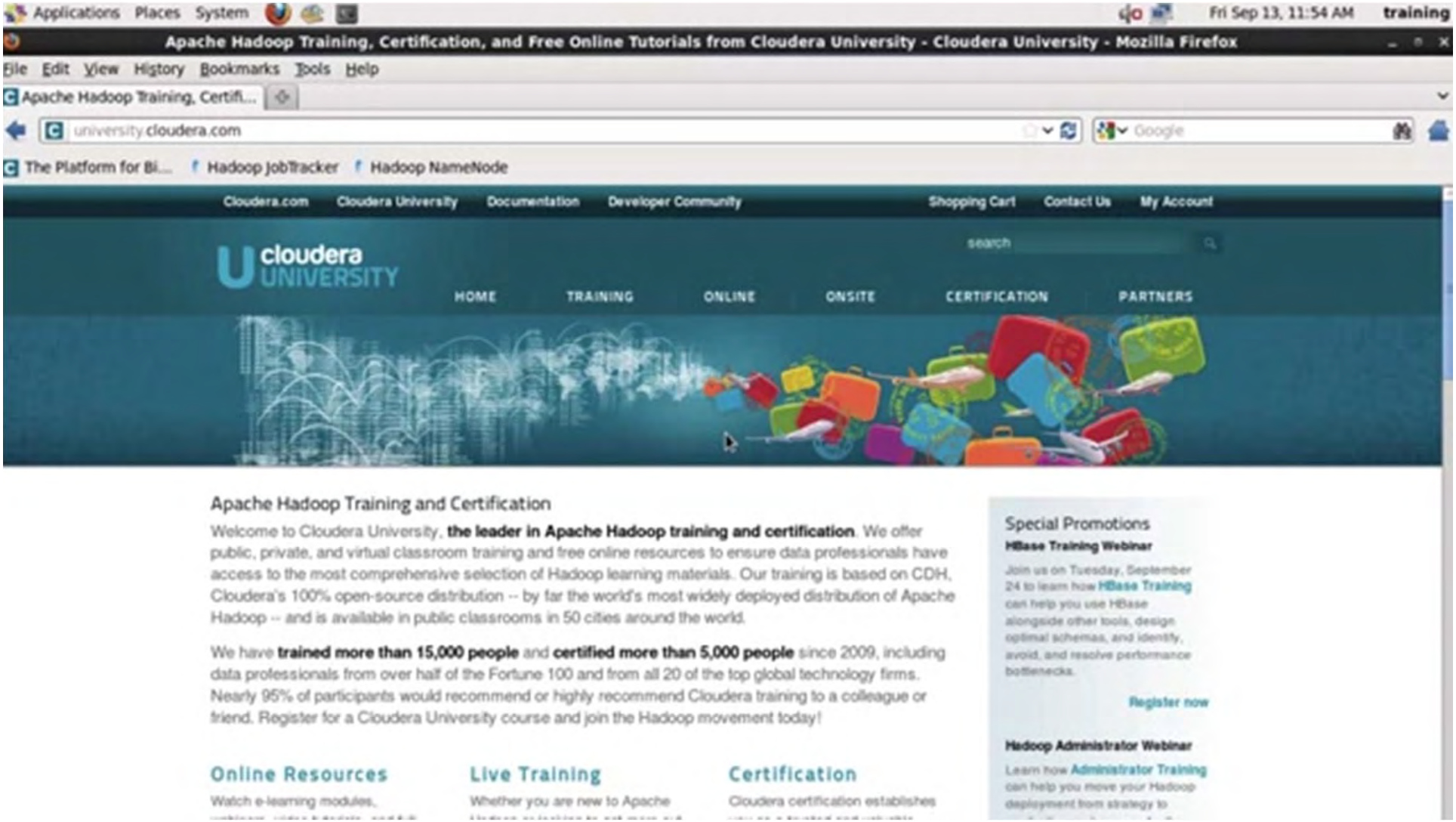

Applications Places System
Fri Sep 13, 11:53 AM training
training@localhost:~/udacity_training/code
File Edit View Search Terminal Help
[training@localhost code]$ ls
mapper.py reducer.py
[training@localhost code]$ ./mapper.py
field1 f2 f3 f4 f5 f6
fieldA fB fC fD fE fF
f3 f5
fC fE
[training@localhost code]$ ls ../data
access_log.gz purchases.txt
[training@localhost code]$ head -50 ../data/purchases.txt > testfile
[training@localhost code]$ cat testfile | ./mapper.py

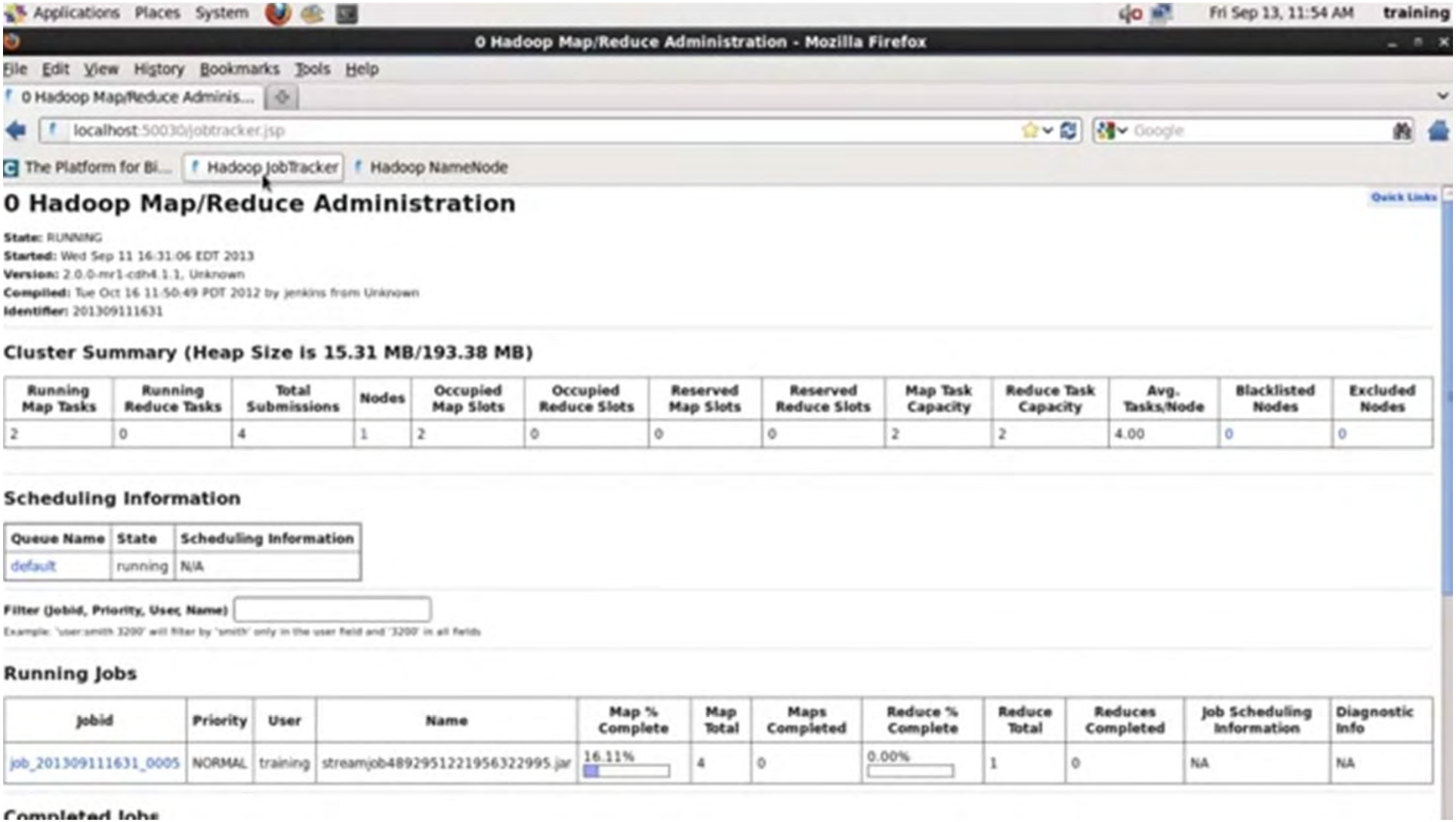

Applications Places System
Fri Sep 13, 11:53 AM training
training@localhost:~/udacity_training/code
File Edit View Search Terminal Help
Riverside 252.88
Tulsa 205.06
Reno 88.25
Chicago 31.08
Fort Wayne 370.55
San Bernardino 170.2
Madison 16.78
Austin 327.75
Portland 108.69
Riverside 15.41
Reno 80.46
Anchorage 298.86
Pittsburgh 475.26
Spokane 3.85
Spokane 287.65
Fresno 466.64
Omaha 255.68
Anchorage 6.38
Aurora 117.81
Philadelphia 351.31
Fremont 222.61
Anchorage 22.36
Norfolk 189.01
Chandler 414.08
Minneapolis 182.05
Honolulu 345.18
Indianapolis 135.96
Chandler 344.09
[training@localhost code]$

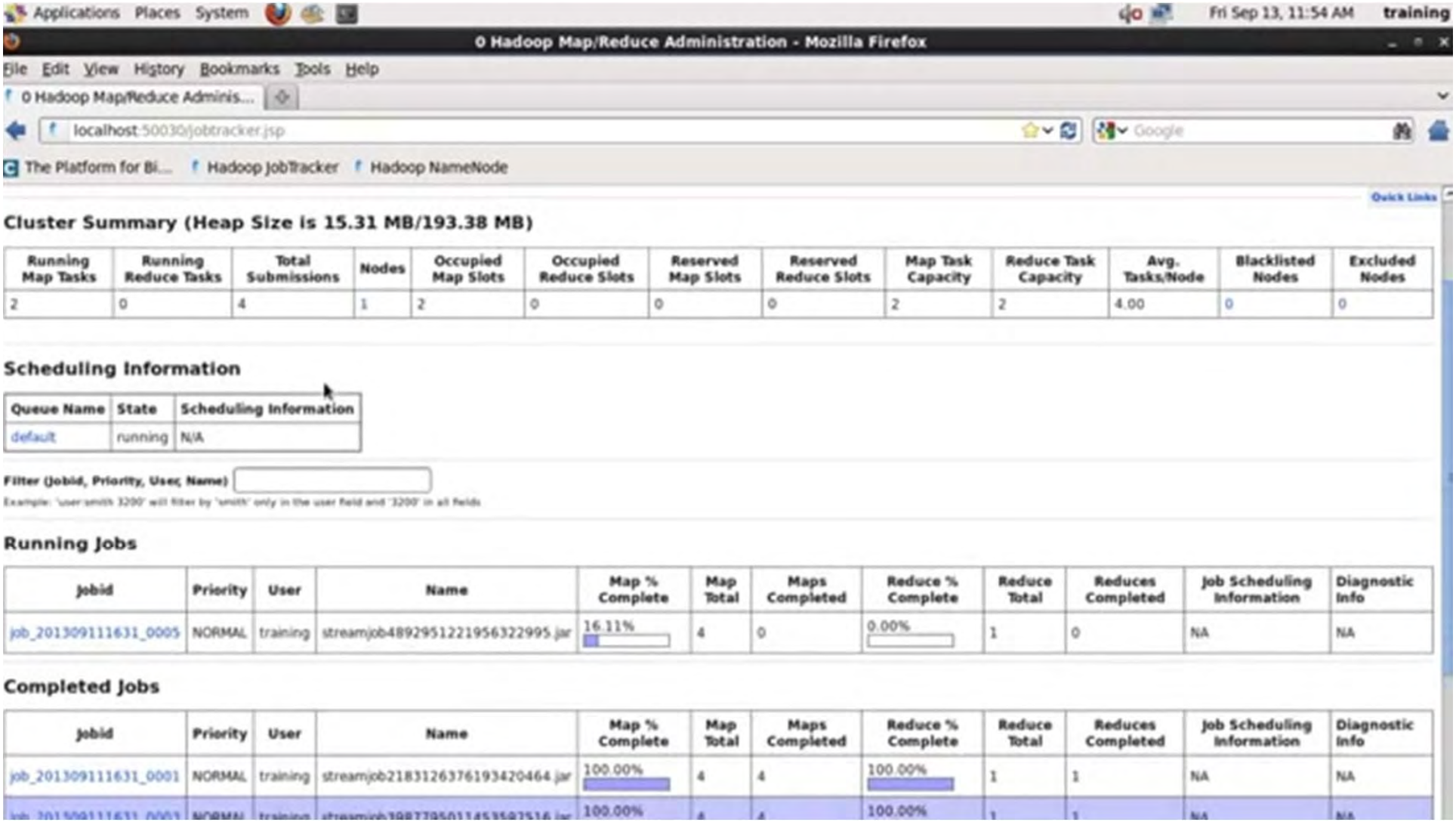

Applications Places System
Fri Sep 13, 11:53 AM training
training@localhost:~/udacity_training/code
File Edit View Search Terminal Help
[training@localhost code]$ cat testfile | ./mapper.py | sort | ./reducer.py

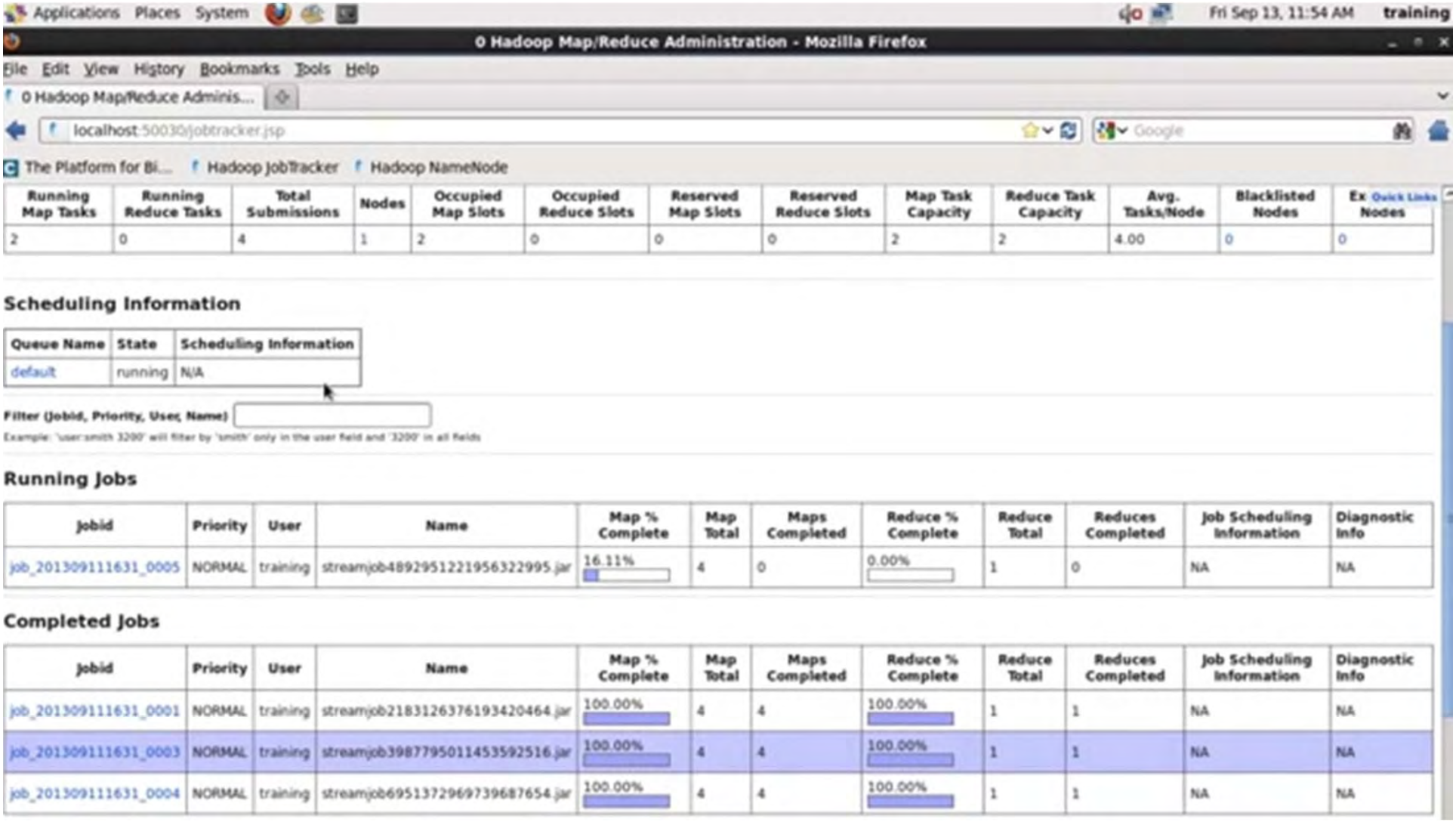

Applications Places System
Fri Sep 13, 11:53 AM training
training@localhost:~/udacity_training/code
File Edit View Search Terminal Help
Fort Worth 367.45
Fremont 222.61
Fresno 466.64
Greensboro 290.82
Honolulu 345.18
Houston 309.16
Indianapolis 135.96
Las Vegas 146.65
Lincoln 136.9
Madison 16.78
Minneapolis 182.05
Newark 39.75
New York 296.8
Norfolk 189.01
Omaha 491.31
Philadelphia 351.31
Pittsburgh 968.77
Portland 108.69
Reno 168.71
Riverside 268.29
San Bernardino 170.2
San Diego 66.08
San Francisco 260.65
San Jose 429.87
Spokane 291.5
Stockton 247.18
Tulsa 205.06
Virginia Beach 376.11
[training@localhost code]$


Applications Places System
Fri Sep 13, 11:54 AM training
training@localhost:~/udacity_training/code
File Edit View Search Terminal Help
[training@localhost code]$ hs mapper.py reducer.py myinput output2
packageJobJar: [mapper.py, reducer.py, /tmp/hadoop-training/hadoop-unjar7183475003578040492/] [] /tmp/stream
job4892951221956322995.jar tmpDir=null
13/09/13 11:54:15 WARN mapred.JobClient: Use GenericOptionsParser for parsing the arguments. Applications sh
ould implement Tool for the same.
13/09/13 11:54:16 WARN snappy.LoadSnappy: Snappy native library is available
13/09/13 11:54:16 INFO snappy.LoadSnappy: Snappy native library loaded
13/09/13 11:54:16 INFO mapred.FileInputFormat: Total input paths to process : 1
13/09/13 11:54:16 INFO streaming.StreamJob: getLocalDirs(): [/var/lib/hadoop-hdfs/cache/training/mapred/loca
l]
13/09/13 11:54:16 INFO streaming.StreamJob: Running job: job_201309111631_0005
13/09/13 11:54:16 INFO streaming.StreamJob: To kill this job, run:
13/09/13 11:54:16 INFO streaming.StreamJ
pred.job.tracker=0.0.0.0:8021 -kill jo
b_201309111631_0005
13/09/13 11:54:16 INFO streaming.StreamJ
.0:50030/jobdetails.jsp?jobid=job_2013
09111631_0005
13/09/13 11:54:17 INFO streaming.StreamJob: map 0% reduce 0%
Apache Hadoop Training, Certification, and Free Onl...


Applications Places System
Fri Sep 13, 11:54 AM training
Apache Hadoop Training, Certification, and Free Online Tutorials from Cloudera University - Cloudera University - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Apache Hadoop Training, Certifi...
university.cloudera.com
Google
The Platform for Bi...
Hadoop JobTracker
Hadoop NameNode
Cloudera.com
Cloudera University
Documentation
Developer Community
Shopping Cart
Contact Us
My Account
search
cloudera UNIVERSITY
HOME
TRAINING
ONLINE
ONSITE
CERTIFICATION
PARTNERS
Apache Hadoop Training and Certification
Welcome to Cloudera University, the leader in Apache Hadoop training and certification. We offer public, private, and virtual classroom training and free online resources to ensure data professionals have access to the most comprehensive selection of Hadoop learning materials. Our training is based on CDH, Cloudera's 100% open-source distribution -- by far the world's most widely deployed distribution of Apache Hadoop -- and is available in public classrooms in 50 cities around the world.
We have trained more than 15,000 people and certified more than 5,000 people since 2009, including data professionals from over half of the Fortune 100 and from all 20 of the top global technology firms. Nearly 95% of participants would recommend or highly recommend Cloudera training to a colleague or friend. Register for a Cloudera University course and join the Hadoop movement today!
Online Resources
Watch e-learning modules,
Live Training
Whether you are new to Apache
Certification
Cloudera certification establishes
Special Promotions
HBase Training Webinar
Join us on Tuesday, September 24 to learn how HBase Training can help you use HBase alongside other tools, design optimal schemas, and identify, avoid, and resolve performance bottlenecks.
Register now
Hadoop Administrator Webinar
Learn how Administrator Training can help you move your Hadoop deployment from strategy to


Applications Places System
Fri Sep 13, 11:54 AM training
0 Hadoop Map/Reduce Administration - Mozilla Firefox
File Edit View History Bookmarks Tools Help
0 Hadoop Map/Reduce Adminis...
localhost:50030/jobtracker.jsp
Google
The Platform for Bi...
Hadoop JobTracker
Hadoop NameNode
0 Hadoop Map/Reduce Administration
Quick Links
State: RUNNING
Started: Wed Sep 11 16:31:06 EDT 2013
Version: 2.0.0-mr1-cdh4.1.1, Unknown
Compiled: Tue Oct 16 11:50:49 PDT 2012 by jenkins from Unknown
Identifier: 201309111631
Cluster Summary (Heap Size is 15.31 MB/193.38 MB)
Running Map Tasks | Running Reduce Tasks | Total Submissions | Nodes | Occupied Map Slots | Occupied Reduce Slots | Reserved Map Slots | Reserved Reduce Slots | Map Task Capacity | Reduce Task Capacity | Avg. Tasks/Node | Blacklisted Nodes | Excluded Nodes
2 | 0 | 4 | 1 | 2 | 0 | 0 | 0 | 2 | 2 | 4.00 | 0 | 0
Scheduling Information
Queue Name | State | Scheduling Information
default | running | N/A
Filter (Jobid, Priority, User, Name)
Example: 'user:smith 3200' will filter by 'smith' only in the user field and '3200' in all fields
Running Jobs
Jobid | Priority | User | Name | Map % Complete | Map Total | Maps Completed | Reduce % Complete | Reduce Total | Reduces Completed | Job Scheduling Information | Diagnostic Info
job_201309111631_0005 | NORMAL | training | streamjob4892951221956322995.jar | 16.11% | 4 | 0 | 0.00% | 1 | 0 | NA | NA
Completed Jobs


Applications Places System
Fri Sep 13, 11:54 AM training
0 Hadoop Map/Reduce Administration - Mozilla Firefox
File Edit View History Bookmarks Tools Help
0 Hadoop Map/Reduce Adminis...
localhost:50030/jobtracker.jsp
Google
The Platform for Bi...
Hadoop JobTracker
Hadoop NameNode
Quick Links
Cluster Summary (Heap Size is 15.31 MB/193.38 MB)
Running Map Tasks | Running Reduce Tasks | Total Submissions | Nodes | Occupied Map Slots | Occupied Reduce Slots | Reserved Map Slots | Reserved Reduce Slots | Map Task Capacity | Reduce Task Capacity | Avg. Tasks/Node | Blacklisted Nodes | Excluded Nodes
2 | 0 | 4 | 1 | 2 | 0 | 0 | 0 | 2 | 2 | 4.00 | 0 | 0
Scheduling Information
Queue Name | State | Scheduling Information
default | running | N/A
Filter (Jobid, Priority, User, Name)
Example: 'user:smith 3200' will filter by 'smith' only in the user field and '3200' in all fields
Running Jobs
Jobid | Priority | User | Name | Map % Complete | Map Total | Maps Completed | Reduce % Complete | Reduce Total | Reduces Completed | Job Scheduling Information | Diagnostic Info
job_201309111631_0005 | NORMAL | training | streamjob4892951221956322995.jar | 16.11% | 4 | 0 | 0.00% | 1 | 0 | NA | NA
Completed Jobs
Jobid | Priority | User | Name | Map % Complete | Map Total | Maps Completed | Reduce % Complete | Reduce Total | Reduces Completed | Job Scheduling Information | Diagnostic Info
job_201309111631_0001 | NORMAL | training | streamjob2183126376193420464.jar | 100.00% | 4 | 4 | 100.00% | 1 | 1 | NA | NA
job_201309111631_0003 | NORMAL | training | streamjob1987795011453592516.jar | 100.00% | 4 | 4 | 100.00% | 1 | 1 | NA | NA

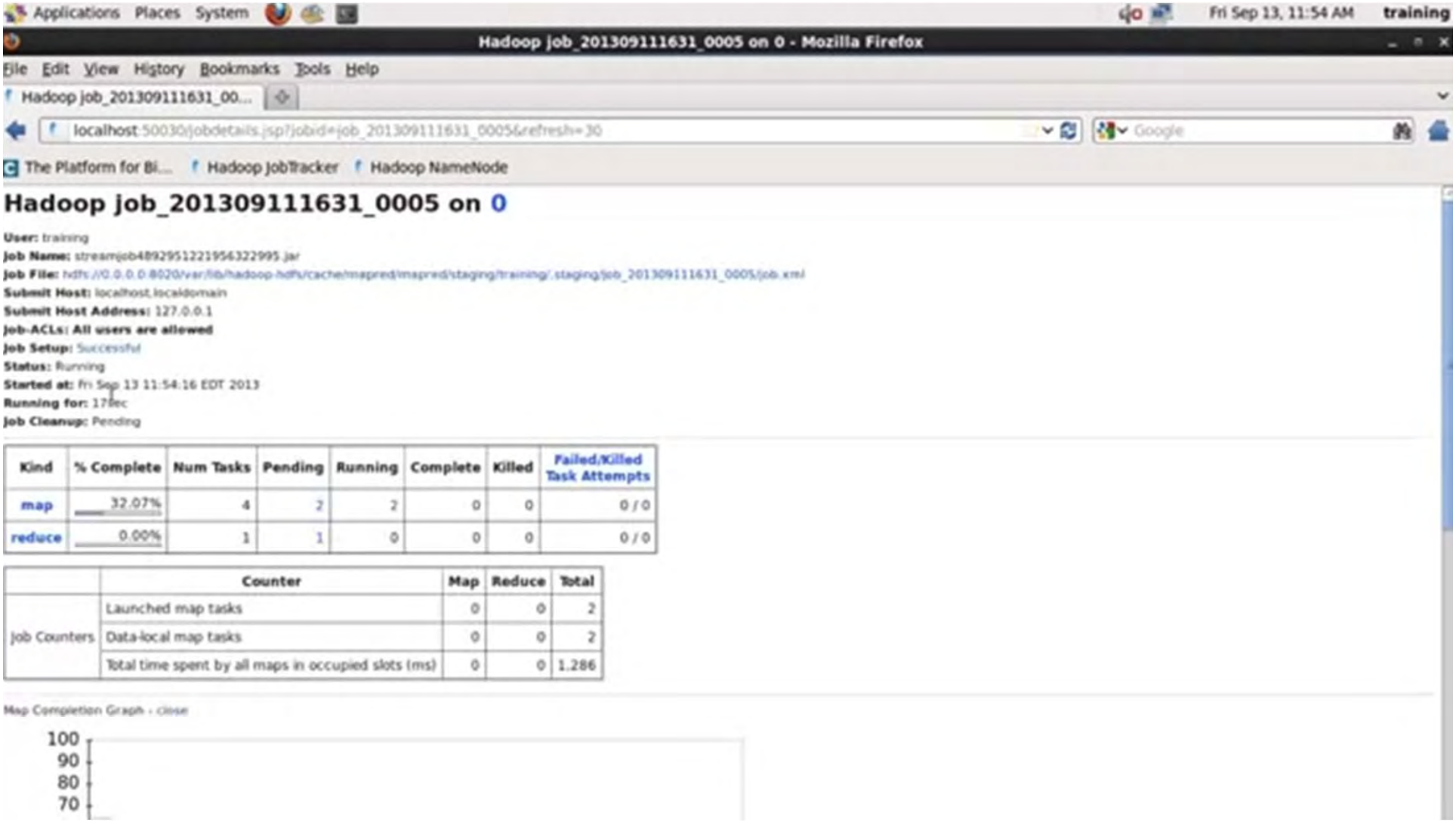


Running Map Tasks	Running Reduce Tasks	Total Submissions	Nodes	Occupied Map Slots	Occupied Reduce Slots	Reserved Map Slots	Reserved Reduce Slots	Map Task Capacity	Reduce Task Capacity	Avg. Tasks/Node	Blacklisted Nodes	Ex Nodes
2	0	4	1	2	0	0	0	2	2	4.00	0	0

Queue Name	State	Scheduling Information
default	running	N/A

Jobid	Priority	User	Name	Map % Complete	Map Total	Maps Completed	Reduce % Complete	Reduce Total	Reduces Completed	Job Scheduling Information	Diagnostic Info
job_201309111631_0005	NORMAL	training	streamjob4892951221956322995.jar	16.11%	4	0	0.00%	1	0	NA	NA

Jobid	Priority	User	Name	Map % Complete	Map Total	Maps Completed	Reduce % Complete	Reduce Total	Reduces Completed	Job Scheduling Information	Diagnostic Info
job_201309111631_0001	NORMAL	training	streamjob2183126376193420464.jar	100.00%	4	4	100.00%	1	1	NA	NA
job_201309111631_0003	NORMAL	training	streamjob3987795011453592516.jar	100.00%	4	4	100.00%	1	1	NA	NA
job_201309111631_0004	NORMAL	training	streamjob6951372969739687654.jar	100.00%	4	4	100.00%	1	1	NA	NA


Applications Places System
Fri Sep 13, 11:54 AM training
Hadoop job_201309111631_0005 on 0 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Hadoop job_201309111631_00...
localhost:50030/jobdetails.jsp?jobid=job_201309111631_0005&refresh=30
Google
The Platform for Bi...
Hadoop JobTracker
Hadoop NameNode
Hadoop job_201309111631_0005 on 0
User: training
Job Name: streamjob4892951221956322995.jar
Job File: hdfs://0.0.0.0:8020/var/lib/hadoop-hdfs/cache/mapred/mapred/staging/training/.staging/job_201309111631_0005/job.xml
Submit Host: localhost.localdomain
Submit Host Address: 127.0.0.1
Job-ACLs: All users are allowed
Job Setup: Successful
Status: Running
Started at: Fri Sep 13 11:54:16 EDT 2013
Running for: 17sec
Job Cleanup: Pending
Kind % Complete Num Tasks Pending Running Complete Killed Failed/Killed Task Attempts
map 32.07% 4 2 2 0 0 0 / 0
reduce 0.00% 1 1 0 0 0 0 / 0
Counter Map Reduce Total
Job Counters Launched map tasks 0 0 2
Data-local map tasks 0 0 2
Total time spent by all maps in occupied slots (ms) 0 0 1,286
Map Completion Graph - close
100
90
80
70

Applications Places System Fri Sep 13, 11:54 AM training

Hadoop map task list for job_201309111631_0005 on 0 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Hadoop job_201309111631_00...

localhost:50030/jobdetails.jsp?jobid=job_201309111631_0005&refresh=30

Google

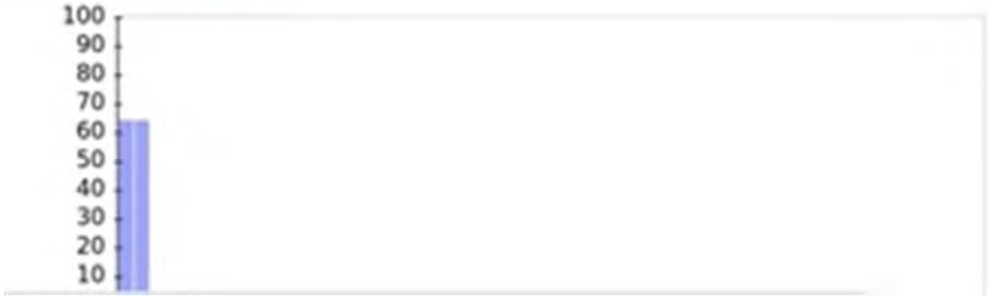
The Platform for Bi... Hadoop JobTracker Hadoop NameNode

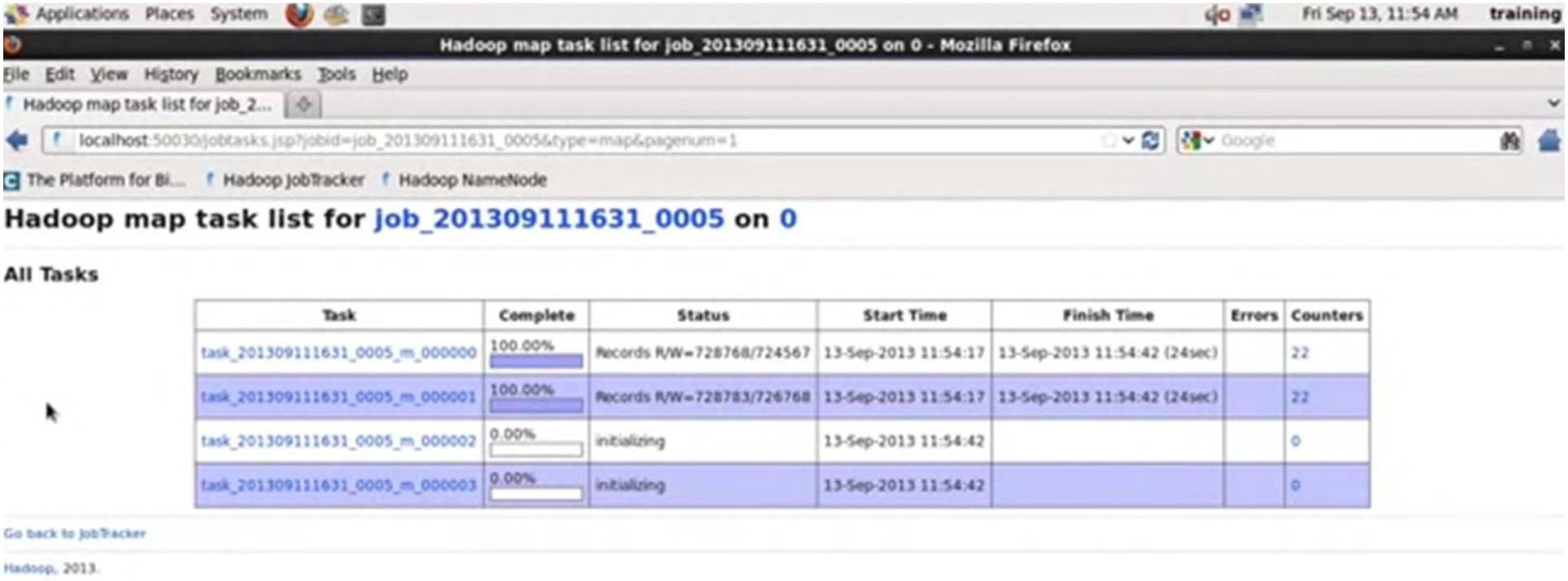
Job-ACLs: All users are allowed
Job Setup: Successful
Status: Running
Started at: Fri Sep 13 11:54:16 EDT 2013
Running for: 17sec
Job Cleanup: Pending

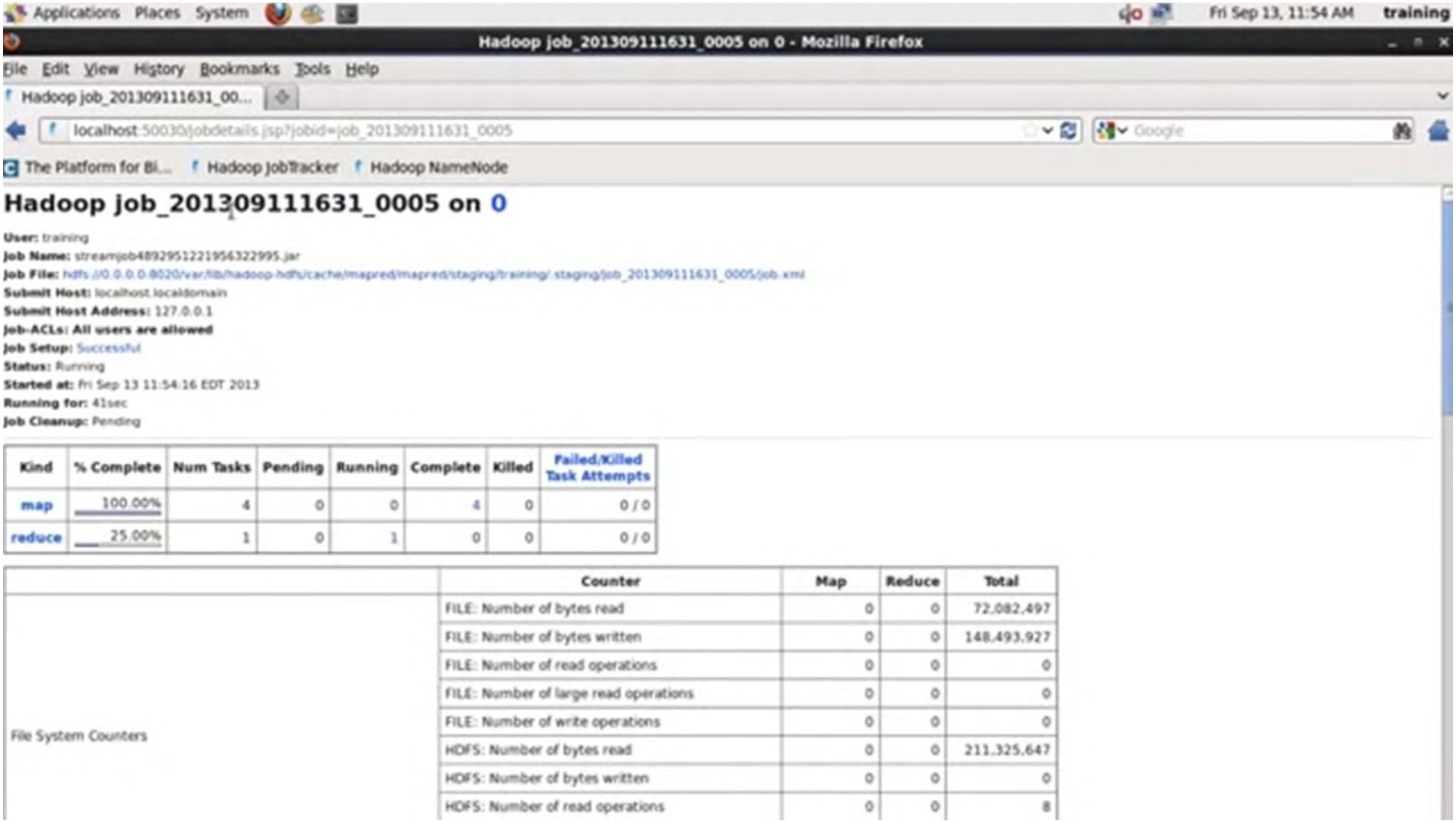
Kind	% Complete	Num Tasks	Pending	Running	Complete	Killed	Failed/Killed Task Attempts
map	32.07%	4	2	2	0	0	0 / 0
reduce	0.00%	1	1	0	0	0	0 / 0

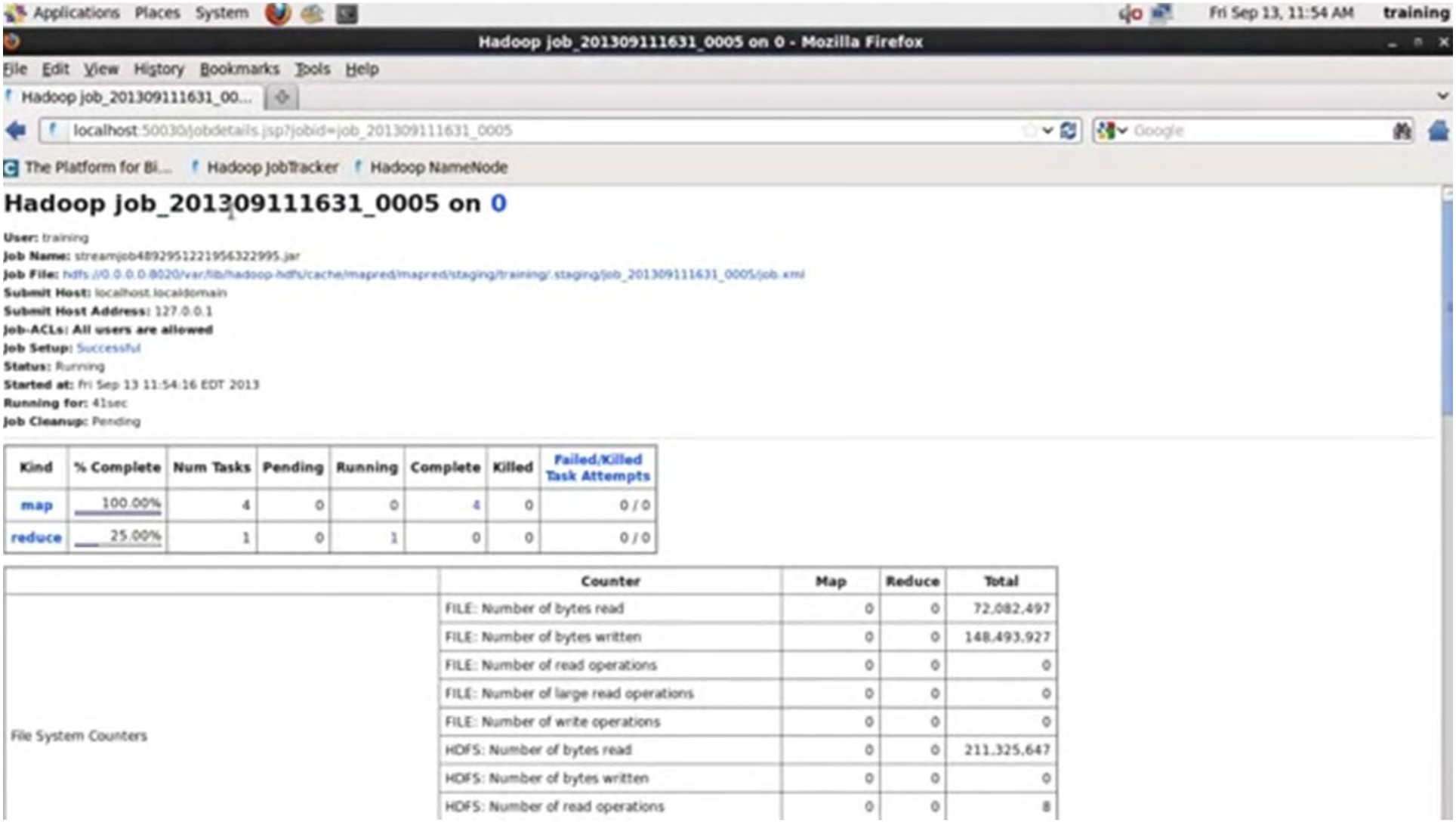
	Counter	Map	Reduce	Total
Job Counters	Launched map tasks	0	0	2
	Data-local map tasks	0	0	2
	Total time spent by all maps in occupied slots (ms)	0	0	1,286

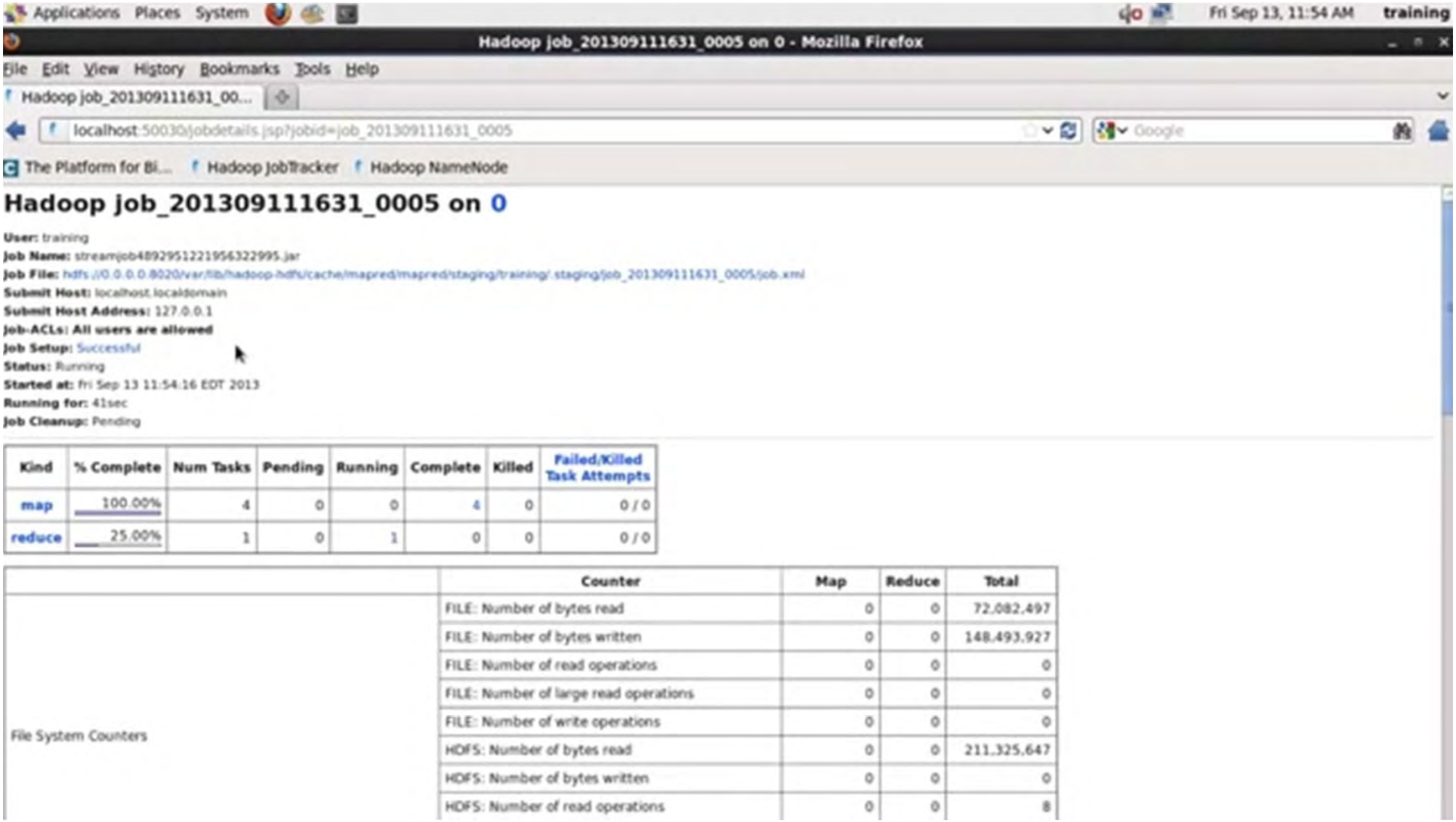
Map Completion Graph - close

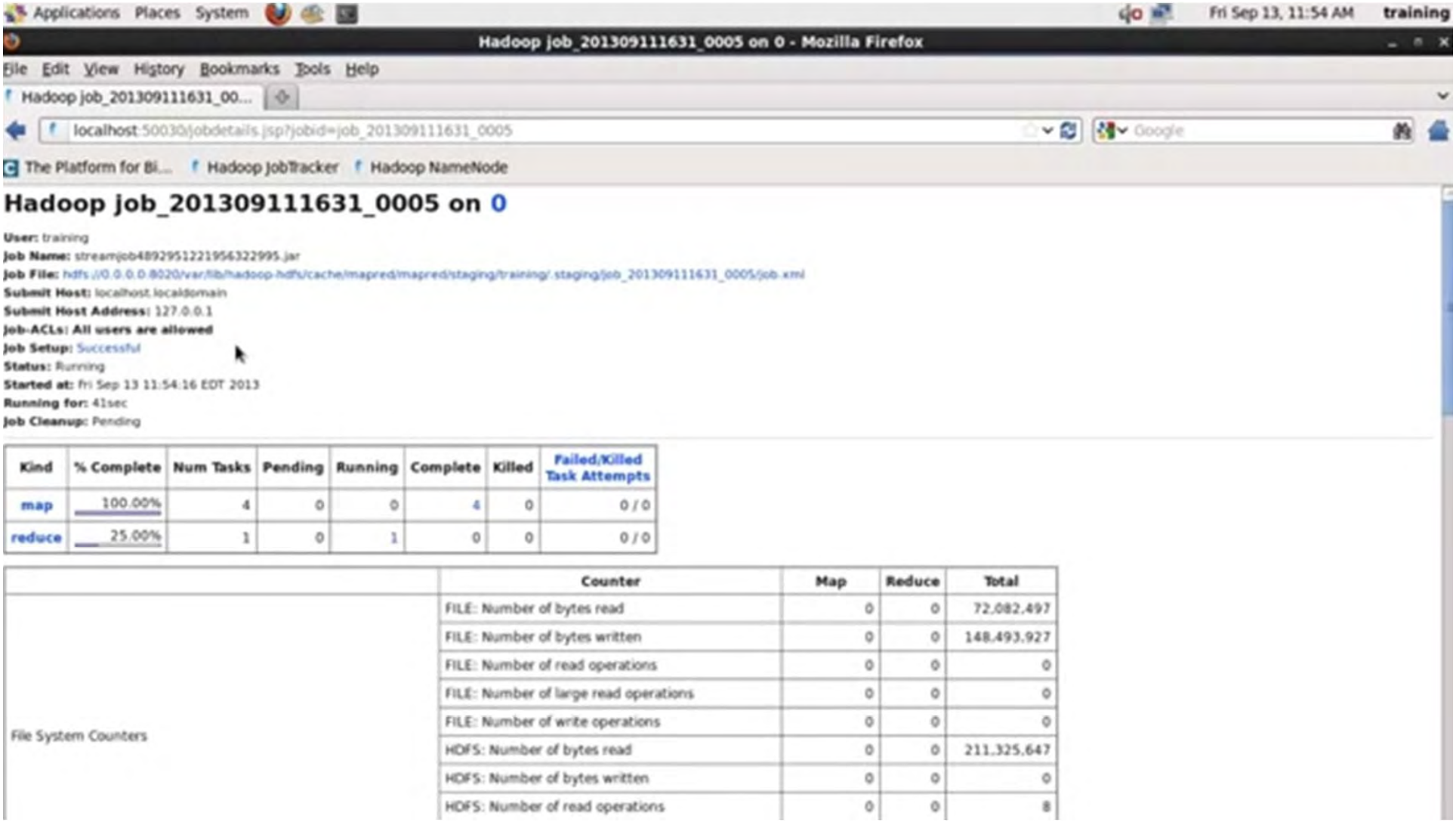


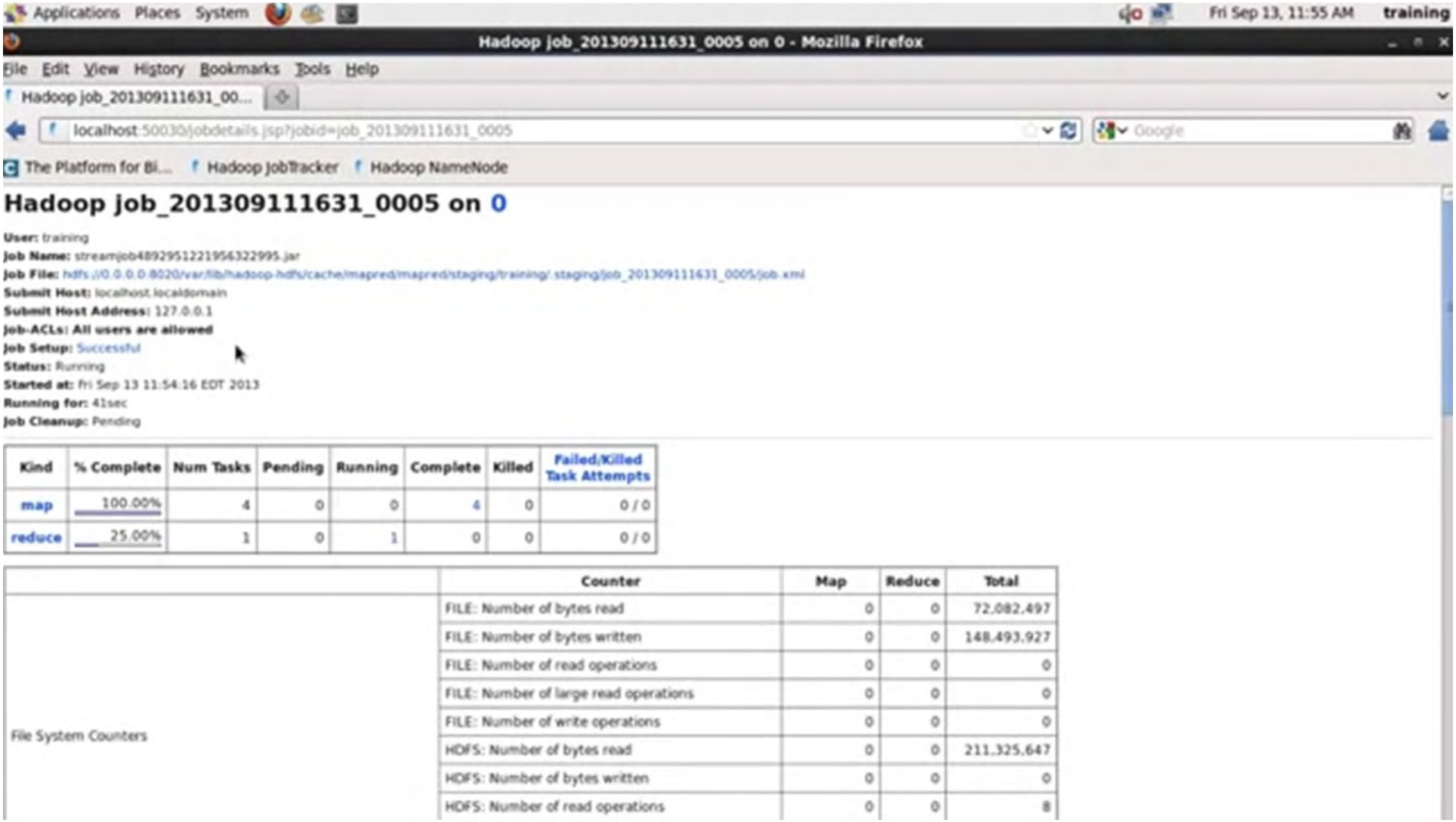

Applications Places System
Fri Sep 13, 11:54 AM training
Hadoop map task list for job_201309111631_0005 on 0 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Hadoop map task list for job_2...
localhost:50030/jobtasks.jsp?jobid=job_201309111631_0005&type=map&pagenum=1
Google
The Platform for Bi...
Hadoop JobTracker
Hadoop NameNode
Hadoop map task list for job_201309111631_0005 on 0
All Tasks
Task
Complete
Status
Start Time
Finish Time
Errors
Counters
task_201309111631_0005_m_000000
100.00%
Records R/W=728768/724567
13-Sep-2013 11:54:17
13-Sep-2013 11:54:42 (24sec)
22
task_201309111631_0005_m_000001
100.00%
Records R/W=728783/726768
13-Sep-2013 11:54:17
13-Sep-2013 11:54:42 (24sec)
22
task_201309111631_0005_m_000002
0.00%
initializing
13-Sep-2013 11:54:42
0
task_201309111631_0005_m_000003
0.00%
initializing
13-Sep-2013 11:54:42
0
Go back to JobTracker
Hadoop, 2013.

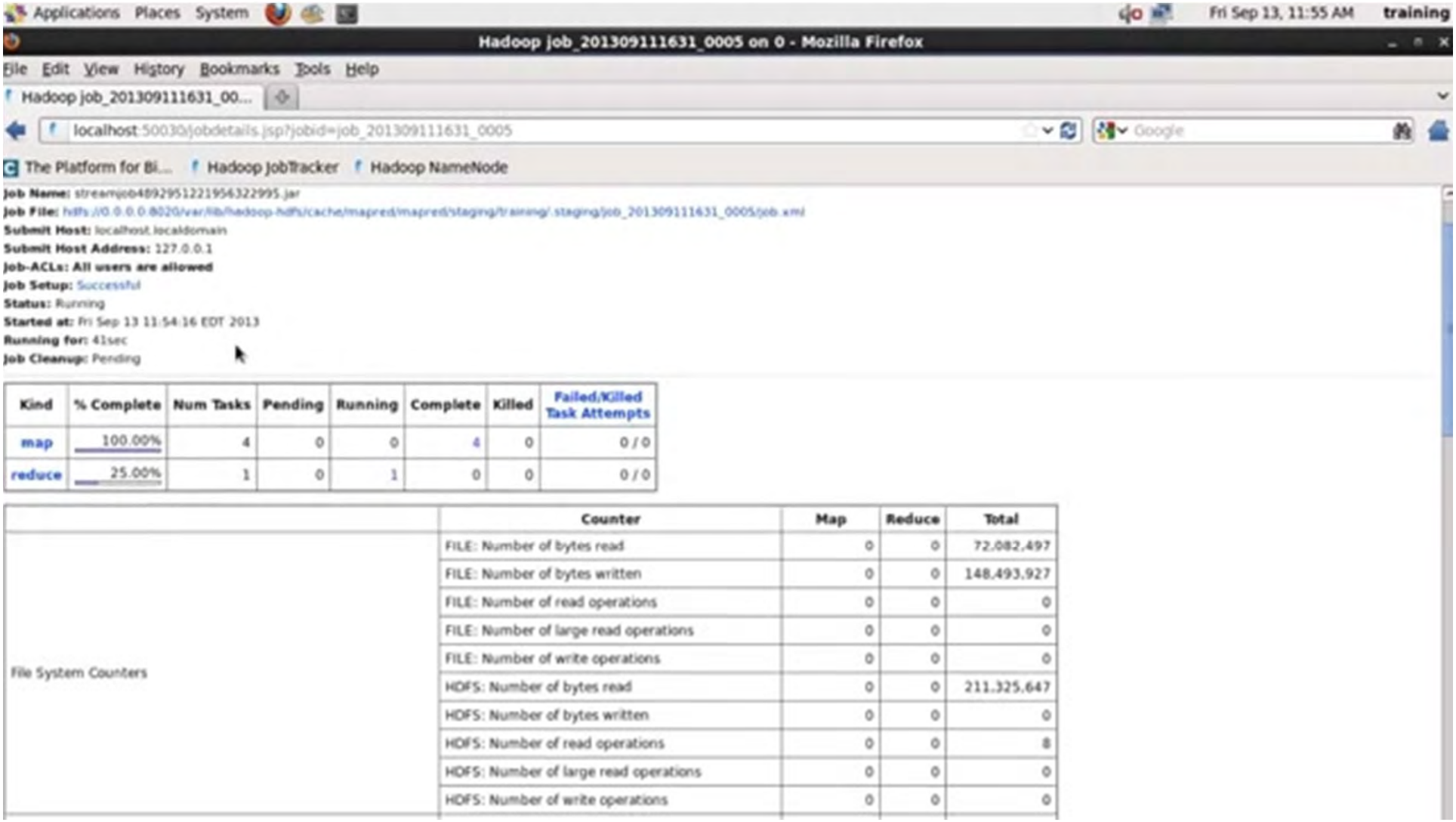

Applications Places System
Fri Sep 13, 11:54 AM training
Hadoop job_201309111631_0005 on 0 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Hadoop job_201309111631_00...
localhost:50030/jobdetails.jsp?jobid=job_201309111631_0005
Google
The Platform for Bi...
Hadoop JobTracker
Hadoop NameNode
Hadoop job_201309111631_0005 on 0
User: training
Job Name: streamjob4892951221956322995.jar
Job File: hdfs://0.0.0.0:8020/var/lib/hadoop-hdfs/cache/mapred/mapred/staging/training/.staging/job_201309111631_0005/job.xml
Submit Host: localhost.localdomain
Submit Host Address: 127.0.0.1
Job-ACLs: All users are allowed
Job Setup: Successful
Status: Running
Started at: Fri Sep 13 11:54:16 EDT 2013
Running for: 41sec
Job Cleanup: Pending
Kind % Complete Num Tasks Pending Running Complete Killed Failed/Killed Task Attempts
map 100.00% 4 0 0 4 0 0 / 0
reduce 25.00% 1 0 1 0 0 0 / 0
Counter Map Reduce Total
File System Counters
FILE: Number of bytes read 0 0 72,082,497
FILE: Number of bytes written 0 0 148,493,927
FILE: Number of read operations 0 0 0
FILE: Number of large read operations 0 0 0
FILE: Number of write operations 0 0 0
HDFS: Number of bytes read 0 0 211,325,647
HDFS: Number of bytes written 0 0 0
HDFS: Number of read operations 0 0 8

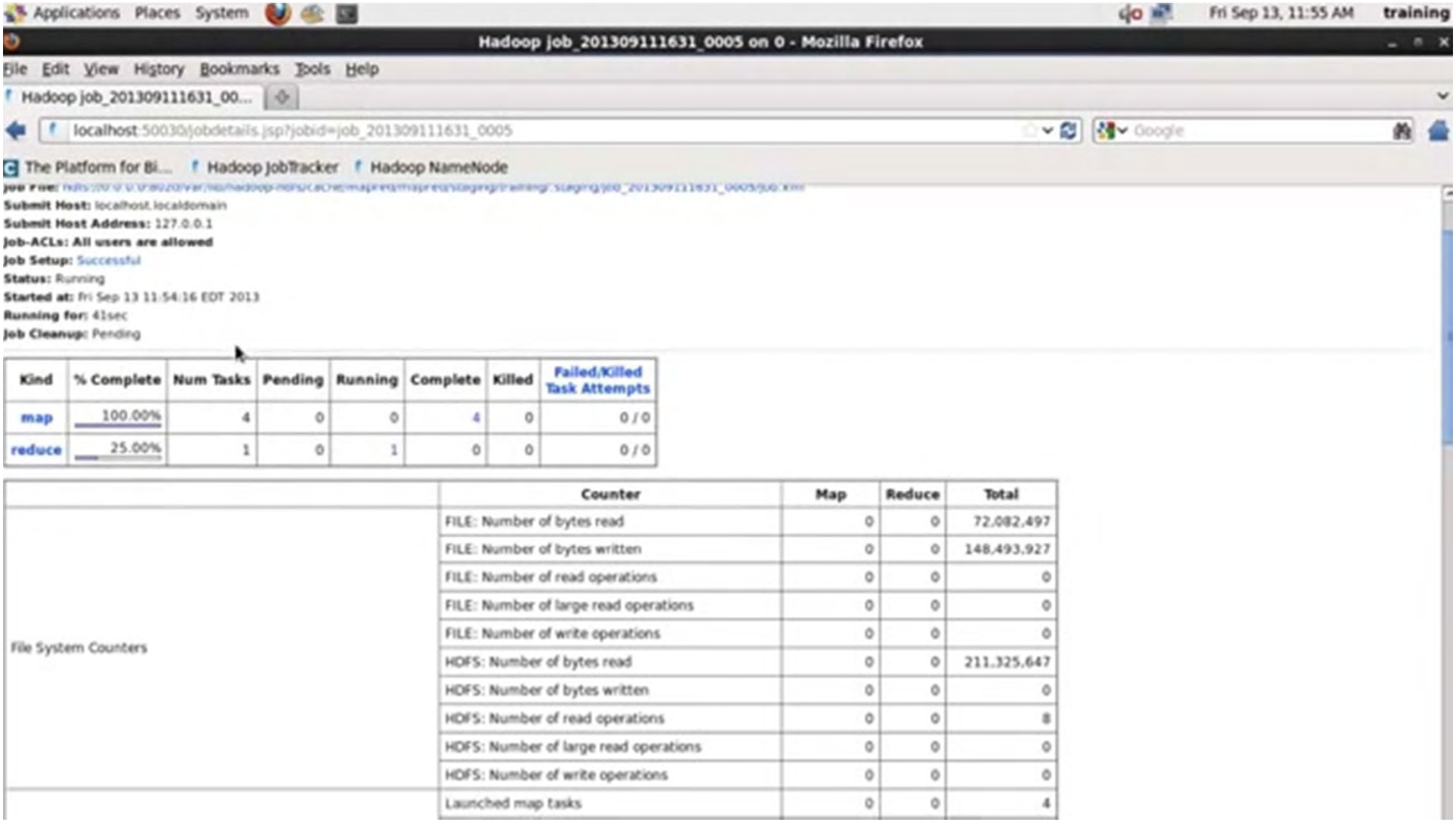

Applications Places System
Fri Sep 13, 11:54 AM training
Hadoop job_201309111631_0005 on 0 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Hadoop job_201309111631_00...
localhost:50030/jobdetails.jsp?jobid=job_201309111631_0005
Google
The Platform for Bi...
Hadoop JobTracker
Hadoop NameNode
Hadoop job_201309111631_0005 on 0
User: training
Job Name: streamjob4892951221956322995.jar
Job File: hdfs://0.0.0.0:8020/var/lib/hadoop-hdfs/cache/mapred/mapred/staging/training/.staging/job_201309111631_0005/job.xml
Submit Host: localhost.localdomain
Submit Host Address: 127.0.0.1
Job-ACLs: All users are allowed
Job Setup: Successful
Status: Running
Started at: Fri Sep 13 11:54:16 EDT 2013
Running for: 41sec
Job Cleanup: Pending
Kind % Complete Num Tasks Pending Running Complete Killed Failed/Killed Task Attempts
map 100.00% 4 0 0 4 0 0 / 0
reduce 25.00% 1 0 1 0 0 0 / 0
Counter Map Reduce Total
File System Counters
FILE: Number of bytes read 0 0 72,082,497
FILE: Number of bytes written 0 0 148,493,927
FILE: Number of read operations 0 0 0
FILE: Number of large read operations 0 0 0
FILE: Number of write operations 0 0 0
HDFS: Number of bytes read 0 0 211,325,647
HDFS: Number of bytes written 0 0 0
HDFS: Number of read operations 0 0 8

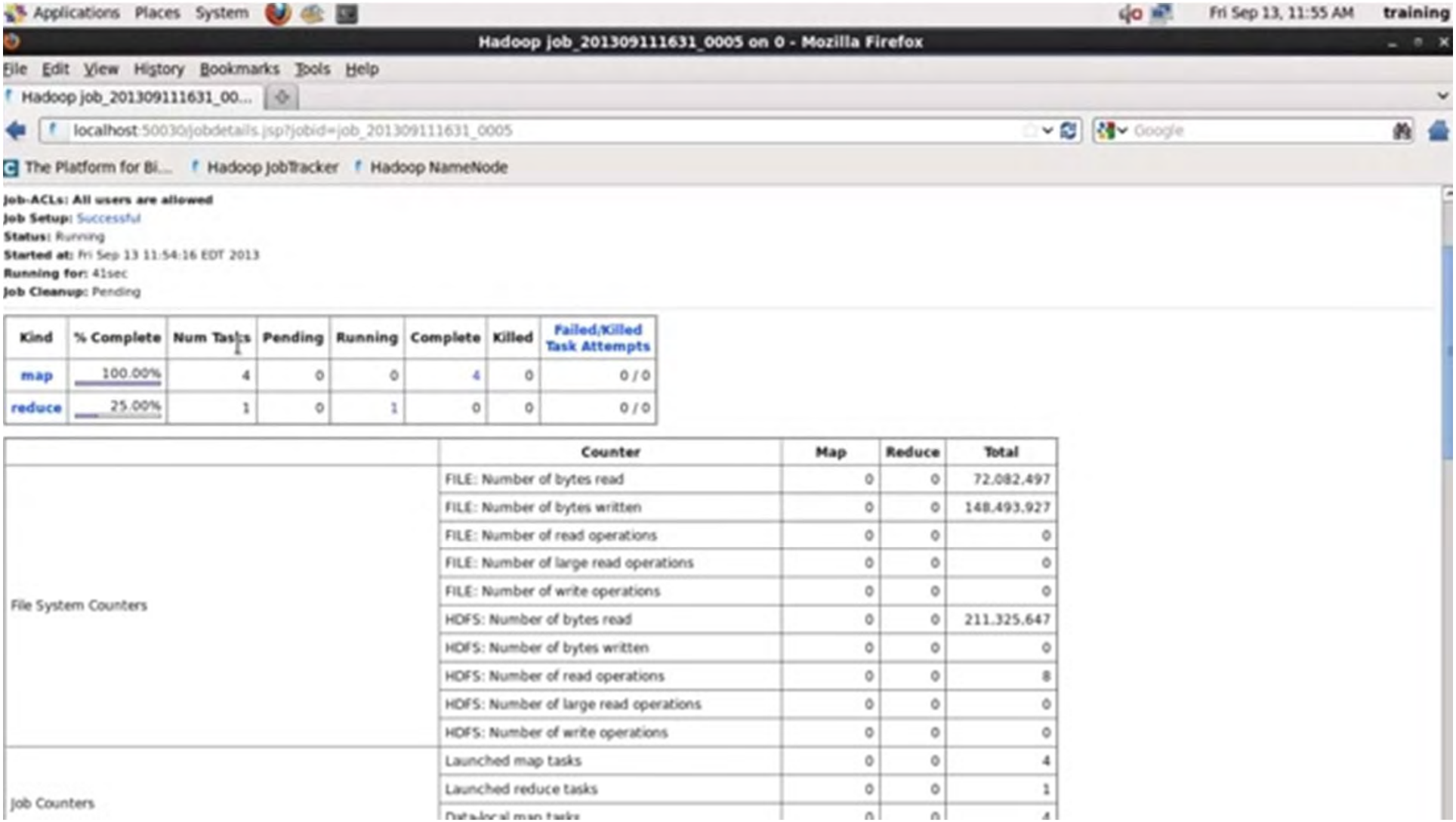

Applications Places System
Fri Sep 13, 11:54 AM training
Hadoop job_201309111631_0005 on 0 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Hadoop job_201309111631_00...
localhost:50030/jobdetails.jsp?jobid=job_201309111631_0005
Google
The Platform for Bi...
Hadoop JobTracker
Hadoop NameNode
Hadoop job_201309111631_0005 on 0
User: training
Job Name: streamjob4892951221956322995.jar
Job File: hdfs://0.0.0.0:8020/var/lib/hadoop-hdfs/cache/mapred/mapred/staging/training/.staging/job_201309111631_0005/job.xml
Submit Host: localhost.localdomain
Submit Host Address: 127.0.0.1
Job-ACLs: All users are allowed
Job Setup: Successful
Status: Running
Started at: Fri Sep 13 11:54:16 EDT 2013
Running for: 41sec
Job Cleanup: Pending
Kind % Complete Num Tasks Pending Running Complete Killed Failed/Killed Task Attempts
map 100.00% 4 0 0 4 0 0 / 0
reduce 25.00% 1 0 1 0 0 0 / 0
Counter Map Reduce Total
File System Counters
FILE: Number of bytes read 0 0 72,082,497
FILE: Number of bytes written 0 0 148,493,927
FILE: Number of read operations 0 0 0
FILE: Number of large read operations 0 0 0
FILE: Number of write operations 0 0 0
HDFS: Number of bytes read 0 0 211,325,647
HDFS: Number of bytes written 0 0 0
HDFS: Number of read operations 0 0 8

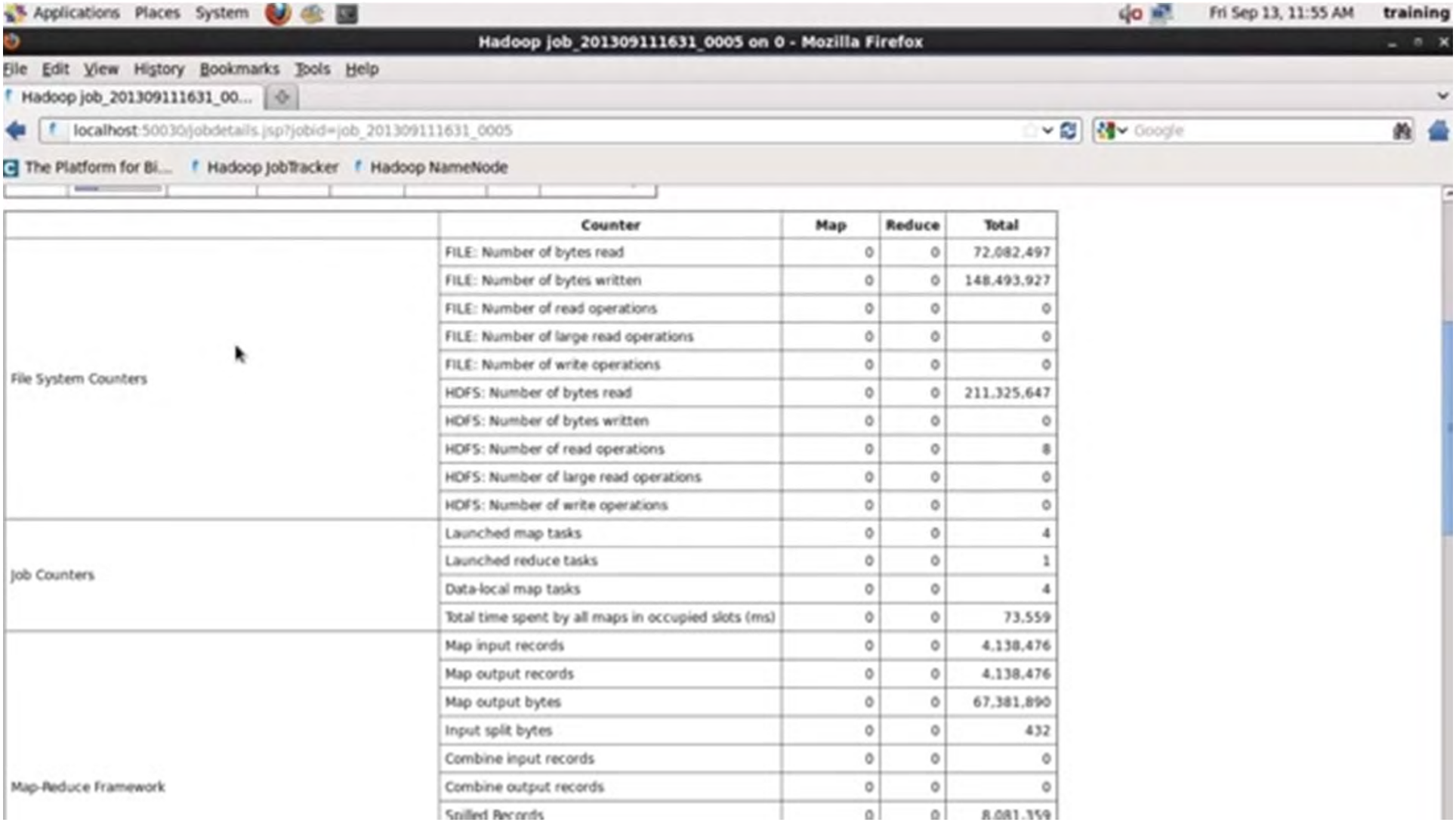

Applications Places System
Fri Sep 13, 11:54 AM training
Hadoop job_201309111631_0005 on 0 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Hadoop job_201309111631_00...
localhost:50030/jobdetails.jsp?jobid=job_201309111631_0005
Google
The Platform for Bi...
Hadoop JobTracker
Hadoop NameNode
Hadoop job_201309111631_0005 on 0
User: training
Job Name: streamjob4892951221956322995.jar
Job File: hdfs://0.0.0.0:8020/var/lib/hadoop-hdfs/cache/mapred/mapred/staging/training/.staging/job_201309111631_0005/job.xml
Submit Host: localhost.localdomain
Submit Host Address: 127.0.0.1
Job-ACLs: All users are allowed
Job Setup: Successful
Status: Running
Started at: Fri Sep 13 11:54:16 EDT 2013
Running for: 41sec
Job Cleanup: Pending
Kind % Complete Num Tasks Pending Running Complete Killed Failed/Killed Task Attempts
map 100.00% 4 0 0 4 0 0 / 0
reduce 25.00% 1 0 1 0 0 0 / 0
Counter Map Reduce Total
File System Counters
FILE: Number of bytes read 0 0 72,082,497
FILE: Number of bytes written 0 0 148,493,927
FILE: Number of read operations 0 0 0
FILE: Number of large read operations 0 0 0
FILE: Number of write operations 0 0 0
HDFS: Number of bytes read 0 0 211,325,647
HDFS: Number of bytes written 0 0 0
HDFS: Number of read operations 0 0 8


Applications Places System
Fri Sep 13, 11:55 AM training
Hadoop job_201309111631_0005 on 0 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Hadoop job_201309111631_00...
localhost:50030/jobdetails.jsp?jobid=job_201309111631_0005
Google
The Platform for Bi...
Hadoop JobTracker
Hadoop NameNode
Hadoop job_201309111631_0005 on 0
User: training
Job Name: streamjob4892951221956322995.jar
Job File: hdfs://0.0.0.0:8020/var/lib/hadoop-hdfs/cache/mapred/mapred/staging/training/.staging/job_201309111631_0005/job.xml
Submit Host: localhost.localdomain
Submit Host Address: 127.0.0.1
Job-ACLs: All users are allowed
Job Setup: Successful
Status: Running
Started at: Fri Sep 13 11:54:16 EDT 2013
Running for: 41sec
Job Cleanup: Pending
Kind % Complete Num Tasks Pending Running Complete Killed Failed/Killed Task Attempts
map 100.00% 4 0 0 4 0 0 / 0
reduce 25.00% 1 0 1 0 0 0 / 0
Counter Map Reduce Total
File System Counters
FILE: Number of bytes read 0 0 72,082,497
FILE: Number of bytes written 0 0 148,493,927
FILE: Number of read operations 0 0 0
FILE: Number of large read operations 0 0 0
FILE: Number of write operations 0 0 0
HDFS: Number of bytes read 0 0 211,325,647
HDFS: Number of bytes written 0 0 0
HDFS: Number of read operations 0 0 8

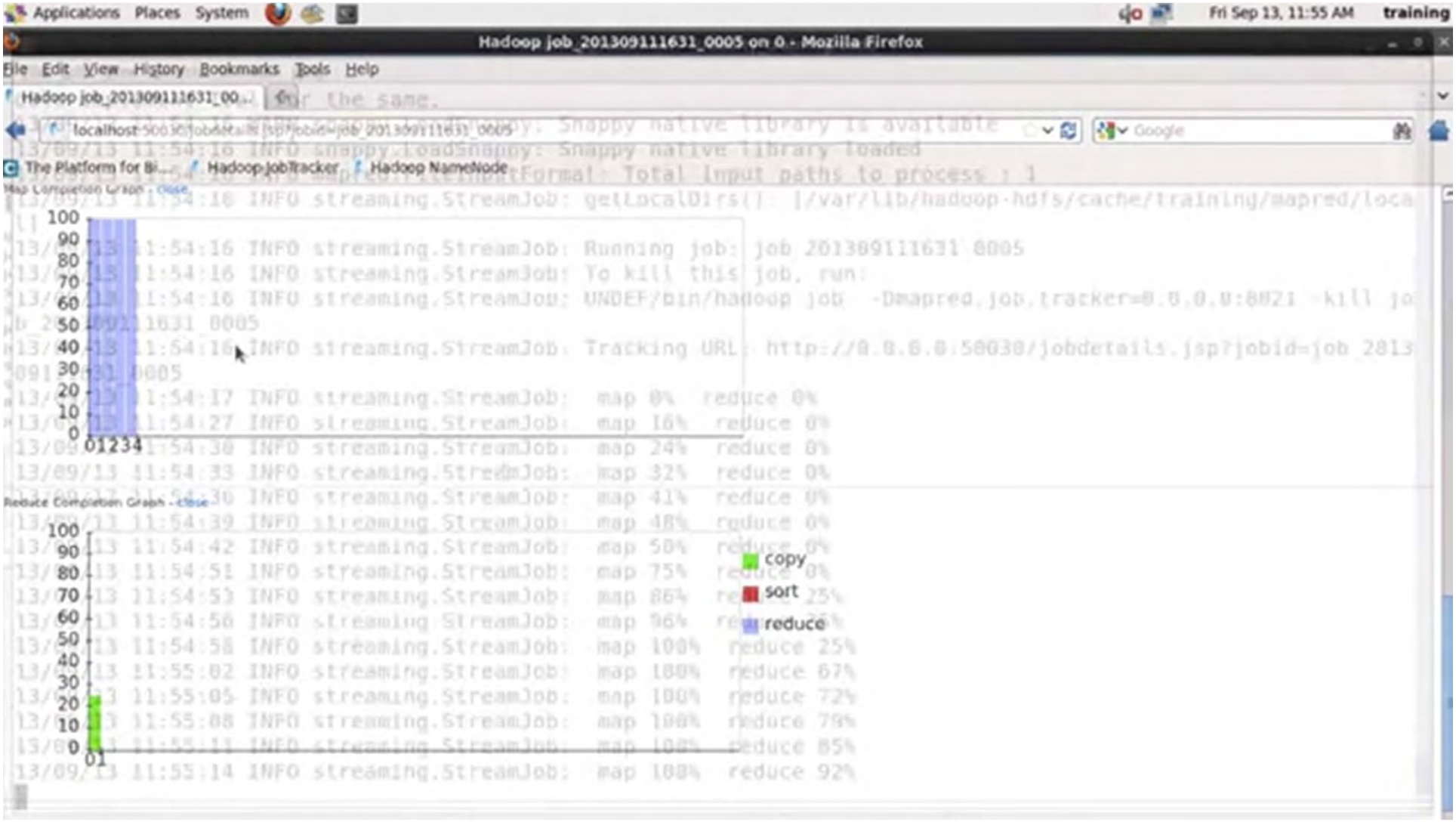

Applications Places System
Fri Sep 13, 11:55 AM training
Hadoop job_201309111631_0005 on 0 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Hadoop job_201309111631_00...
localhost:50030/jobdetails.jsp?jobid=job_201309111631_0005
Google
The Platform for Bi...
Hadoop JobTracker
Hadoop NameNode
Job Name: streamjob4892951221956322995.jar
Job File: hdfs://0.0.0.0:8020/var/lib/hadoop-hdfs/cache/mapred/mapred/staging/training/.staging/job_201309111631_0005/job.xml
Submit Host: localhost.localdomain
Submit Host Address: 127.0.0.1
Job-ACLs: All users are allowed
Job Setup: Successful
Status: Running
Started at: Fri Sep 13 11:54:16 EDT 2013
Running for: 41sec
Job Cleanup: Pending
Kind % Complete Num Tasks Pending Running Complete Killed Failed/Killed Task Attempts
map 100.00% 4 0 0 4 0 0 / 0
reduce 25.00% 1 0 1 0 0 0 / 0
Counter Map Reduce Total
File System Counters
FILE: Number of bytes read 0 0 72,082,497
FILE: Number of bytes written 0 0 148,493,927
FILE: Number of read operations 0 0 0
FILE: Number of large read operations 0 0 0
FILE: Number of write operations 0 0 0
HDFS: Number of bytes read 0 0 211,325,647
HDFS: Number of bytes written 0 0 0
HDFS: Number of read operations 0 0 8
HDFS: Number of large read operations 0 0 0
HDFS: Number of write operations 0 0 0

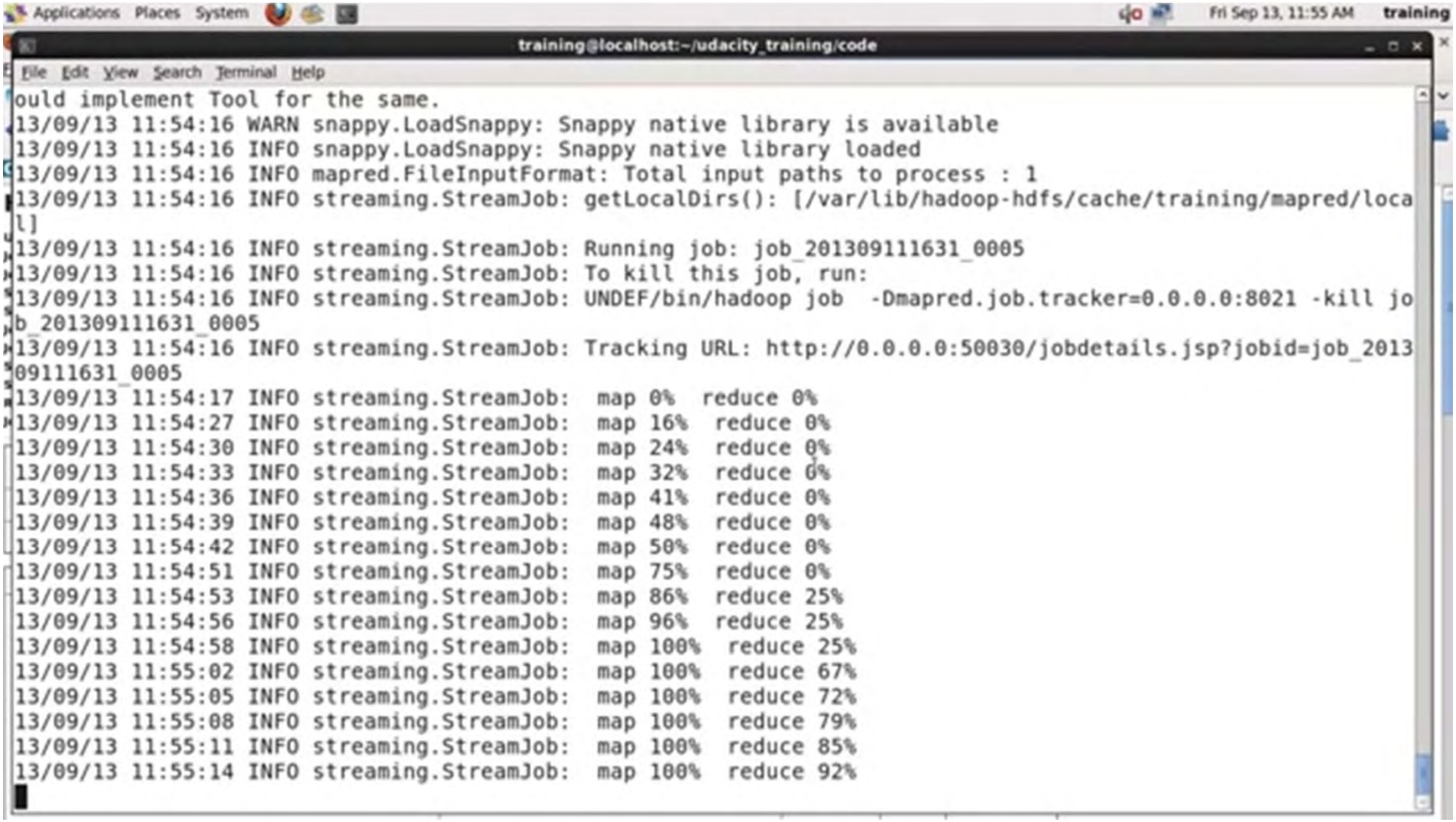

Applications Places System
Fri Sep 13, 11:55 AM training
Hadoop job_201309111631_0005 on 0 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Hadoop job_201309111631_00...
localhost:50030/jobdetails.jsp?jobid=job_201309111631_0005
Google
The Platform for Bi...
Hadoop JobTracker
Hadoop NameNode
Submit Host: localhost.localdomain
Submit Host Address: 127.0.0.1
Job-ACLs: All users are allowed
Job Setup: Successful
Status: Running
Started at: Fri Sep 13 11:54:16 EDT 2013
Running for: 41sec
Job Cleanup: Pending
Kind % Complete Num Tasks Pending Running Complete Killed Failed/Killed Task Attempts
map 100.00% 4 0 0 4 0 0 / 0
reduce 25.00% 1 0 1 0 0 0 / 0
Counter Map Reduce Total
File System Counters
FILE: Number of bytes read 0 0 72,082,497
FILE: Number of bytes written 0 0 148,493,927
FILE: Number of read operations 0 0 0
FILE: Number of large read operations 0 0 0
FILE: Number of write operations 0 0 0
HDFS: Number of bytes read 0 0 211,325,647
HDFS: Number of bytes written 0 0 0
HDFS: Number of read operations 0 0 8
HDFS: Number of large read operations 0 0 0
HDFS: Number of write operations 0 0 0
Launched map tasks 0 0 4

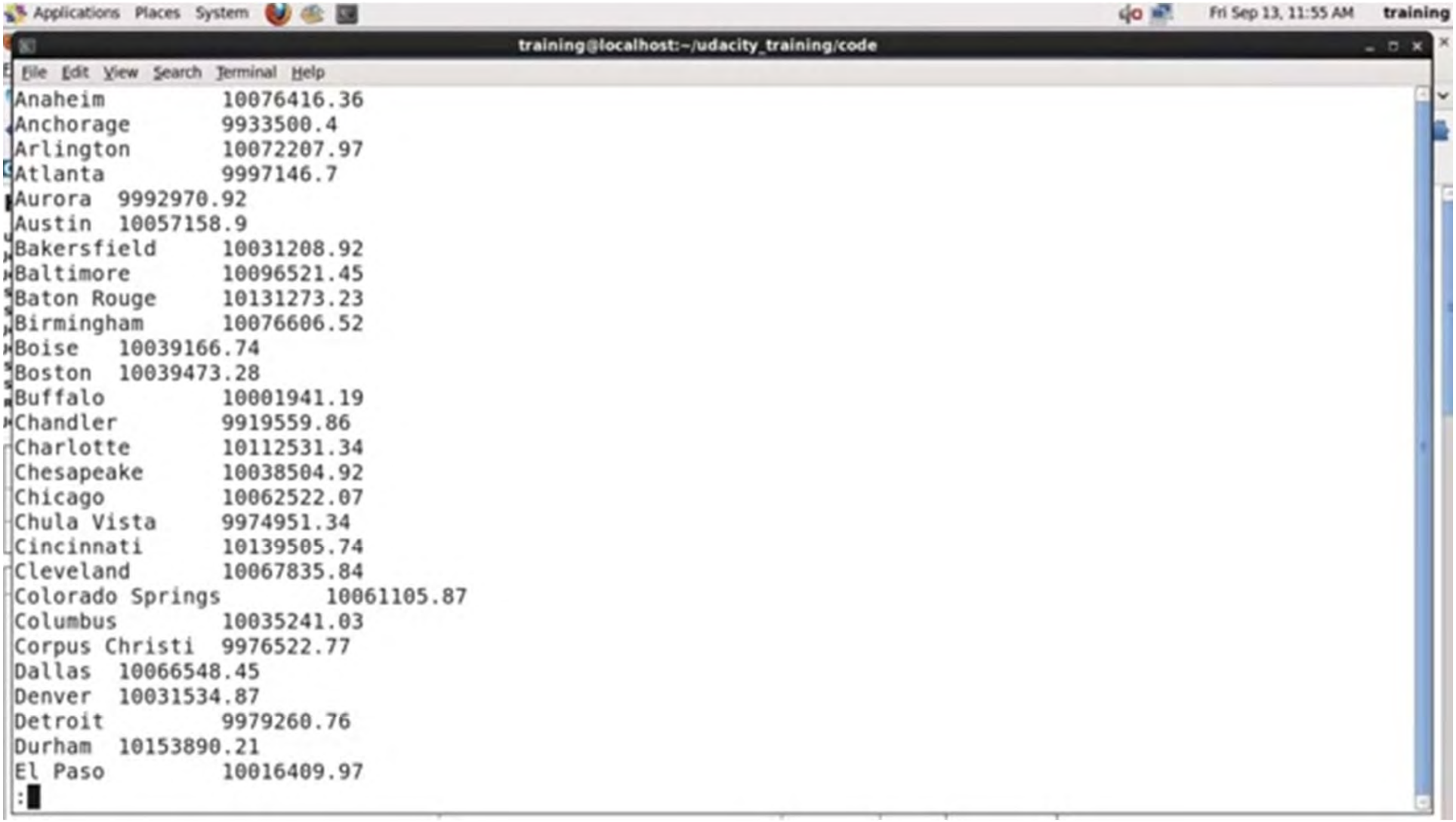

Applications Places System
Fri Sep 13, 11:55 AM training
Hadoop job_201309111631_0005 on 0 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Hadoop job_201309111631_00...
localhost:50030/jobdetails.jsp?jobid=job_201309111631_0005
Google
The Platform for Bi...
Hadoop JobTracker
Hadoop NameNode
Job-ACLs: All users are allowed
Job Setup: Successful
Status: Running
Started at: Fri Sep 13 11:54:16 EDT 2013
Running for: 41sec
Job Cleanup: Pending
Kind % Complete Num Tasks Pending Running Complete Killed Failed/Killed Task Attempts
map 100.00% 4 0 0 4 0 0 / 0
reduce 25.00% 1 0 1 0 0 0 / 0
Counter Map Reduce Total
File System Counters
FILE: Number of bytes read 0 0 72,082,497
FILE: Number of bytes written 0 0 148,493,927
FILE: Number of read operations 0 0 0
FILE: Number of large read operations 0 0 0
FILE: Number of write operations 0 0 0
HDFS: Number of bytes read 0 0 211,325,647
HDFS: Number of bytes written 0 0 0
HDFS: Number of read operations 0 0 8
HDFS: Number of large read operations 0 0 0
HDFS: Number of write operations 0 0 0
Job Counters
Launched map tasks 0 0 4
Launched reduce tasks 0 0 1



	Counter	Map	Reduce	Total
File System Counters	FILE: Number of bytes read	0	0	72,082,497
	FILE: Number of bytes written	0	0	148,493,927
	FILE: Number of read operations	0	0	0
	FILE: Number of large read operations	0	0	0
	FILE: Number of write operations	0	0	0
	HDFS: Number of bytes read	0	0	211,325,647
	HDFS: Number of bytes written	0	0	0
	HDFS: Number of read operations	0	0	8
	HDFS: Number of large read operations	0	0	0
	HDFS: Number of write operations	0	0	0
Job Counters	Launched map tasks	0	0	4
	Launched reduce tasks	0	0	1
	Data-local map tasks	0	0	4
	Total time spent by all maps in occupied slots (ms)	0	0	73,559
Map-Reduce Framework	Map input records	0	0	4,138,476
	Map output records	0	0	4,138,476
	Map output bytes	0	0	67,381,890
	Input split bytes	0	0	432
	Combine input records	0	0	0
	Combine output records	0	0	0
	Spilled Records	0	0	8,081,359


Applications Places System
Fri Sep 13, 11:55 AM training
Hadoop job_201309111631_0005 on 0 - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Hadoop job_201309111631_00...
localhost:50030/jobdetails.jsp?jobid=job_201309111631_0005
Google
The Platform for Bi...
Hadoop JobTracker
Hadoop NameNode
Map Completion Graph - close
100
90
80
70
60
50
40
30
20
10
0
01234
Reduce Completion Graph - close
100
90
80
70
60
50
40
30
20
10
0
01
copy
sort
reduce


Applications Places System
Fri Sep 13, 11:55 AM training
training@localhost:~/udacity_training/code
File Edit View Search Terminal Help
ould implement Tool for the same.
13/09/13 11:54:16 WARN snappy.LoadSnappy: Snappy native library is available
13/09/13 11:54:16 INFO snappy.LoadSnappy: Snappy native library loaded
13/09/13 11:54:16 INFO mapred.FileInputFormat: Total input paths to process : 1
13/09/13 11:54:16 INFO streaming.StreamJob: getLocalDirs(): [/var/lib/hadoop-hdfs/cache/training/mapred/loca
l]
13/09/13 11:54:16 INFO streaming.StreamJob: Running job: job_201309111631_0005
13/09/13 11:54:16 INFO streaming.StreamJob: To kill this job, run:
13/09/13 11:54:16 INFO streaming.StreamJob: UNDEF/bin/hadoop job -Dmapred.job.tracker=0.0.0.0:8021 -kill jo
b_201309111631_0005
13/09/13 11:54:16 INFO streaming.StreamJob: Tracking URL: http://0.0.0.0:50030/jobdetails.jsp?jobid=job_2013
09111631_0005
13/09/13 11:54:17 INFO streaming.StreamJob: map 0% reduce 0%
13/09/13 11:54:27 INFO streaming.StreamJob: map 16% reduce 0%
13/09/13 11:54:30 INFO streaming.StreamJob: map 24% reduce 0%
13/09/13 11:54:33 INFO streaming.StreamJob: map 32% reduce 0%
13/09/13 11:54:36 INFO streaming.StreamJob: map 41% reduce 0%
13/09/13 11:54:39 INFO streaming.StreamJob: map 48% reduce 0%
13/09/13 11:54:42 INFO streaming.StreamJob: map 50% reduce 0%
13/09/13 11:54:51 INFO streaming.StreamJob: map 75% reduce 0%
13/09/13 11:54:53 INFO streaming.StreamJob: map 86% reduce 25%
13/09/13 11:54:56 INFO streaming.StreamJob: map 96% reduce 25%
13/09/13 11:54:58 INFO streaming.StreamJob: map 100% reduce 25%
13/09/13 11:55:02 INFO streaming.StreamJob: map 100% reduce 67%
13/09/13 11:55:05 INFO streaming.StreamJob: map 100% reduce 72%
13/09/13 11:55:08 INFO streaming.StreamJob: map 100% reduce 79%
13/09/13 11:55:11 INFO streaming.StreamJob: map 100% reduce 85%
13/09/13 11:55:14 INFO streaming.StreamJob: map 100% reduce 92%


Applications Places System
Fri Sep 13, 11:55 AM training
training@localhost:~/udacity_training/code
File Edit View Search Terminal Help
Anaheim 10076416.36
Anchorage 9933500.4
Arlington 10072207.97
Atlanta 9997146.7
Aurora 9992970.92
Austin 10057158.9
Bakersfield 10031208.92
Baltimore 10096521.45
Baton Rouge 10131273.23
Birmingham 10076606.52
Boise 10039166.74
Boston 10039473.28
Buffalo 10001941.19
Chandler 9919559.86
Charlotte 10112531.34
Chesapeake 10038504.92
Chicago 10062522.07
Chula Vista 9974951.34
Cincinnati 10139505.74
Cleveland 10067835.84
Colorado Springs 10061105.87
Columbus 10035241.03
Corpus Christi 9976522.77
Dallas 10066548.45
Denver 10031534.87
Detroit 9979260.76
Durham 10153890.21
El Paso 10016409.97
: