Gaurab Dey
Email: gaurabde@ufl.edu

Homework-02

UFID: 78029140

Answer-1: Screen-shots of various WebUI of hadoop-cluster

Figure-1: JobTracker portal

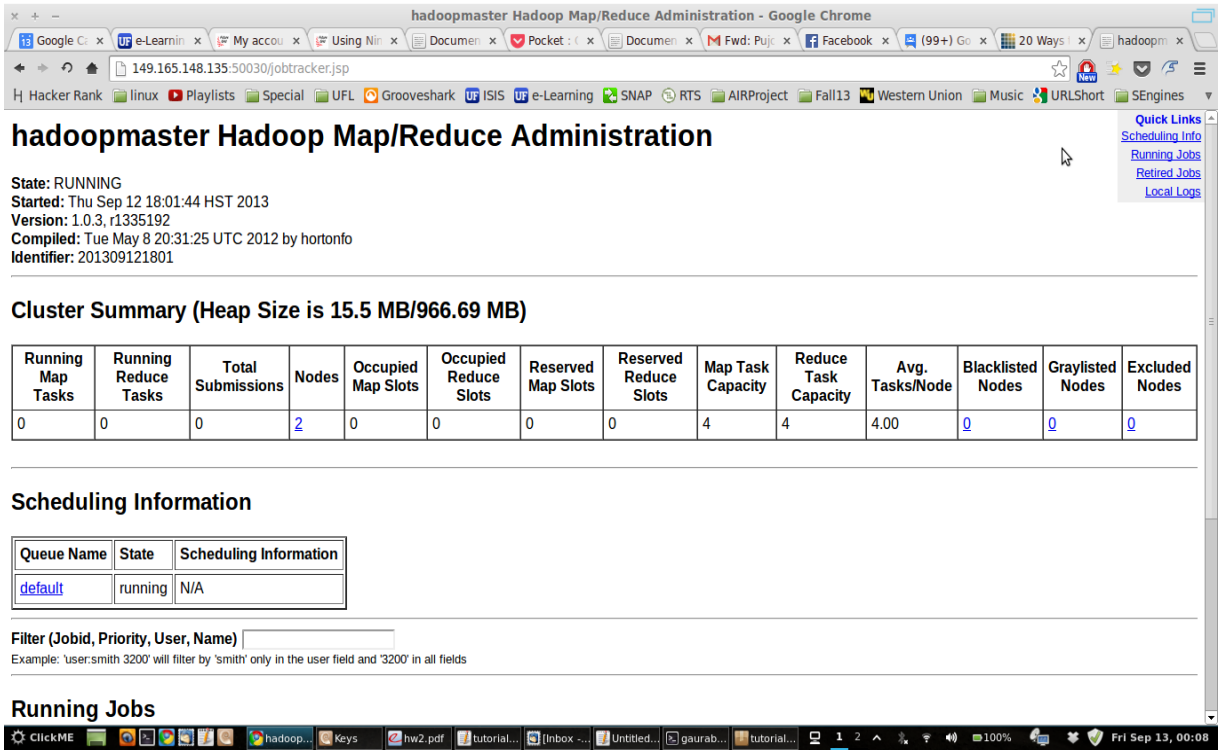
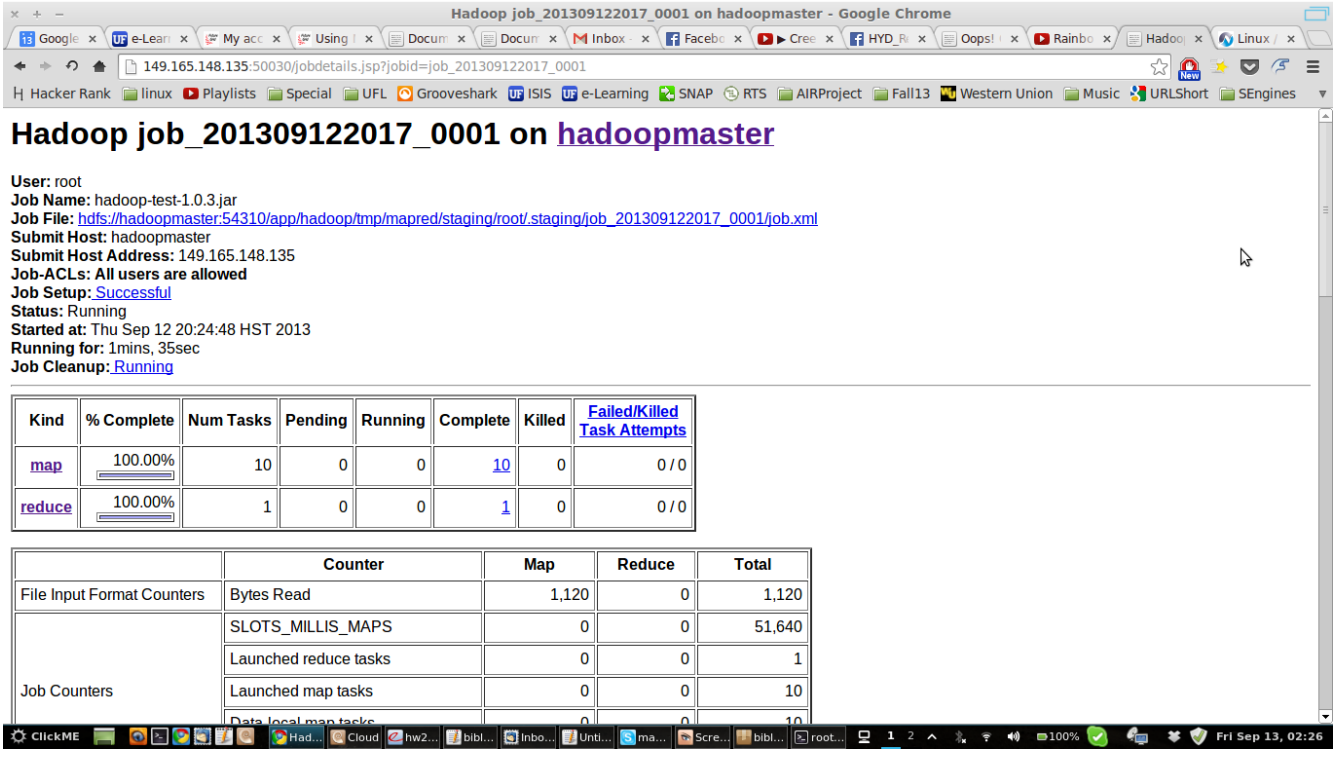


Figure-2: Map-Reduce Portal



Gaurab Dey
Email: gaurabde@ufl.edu

UFID: 78029140

Figure-3: Map-Reduce Completion Graphs

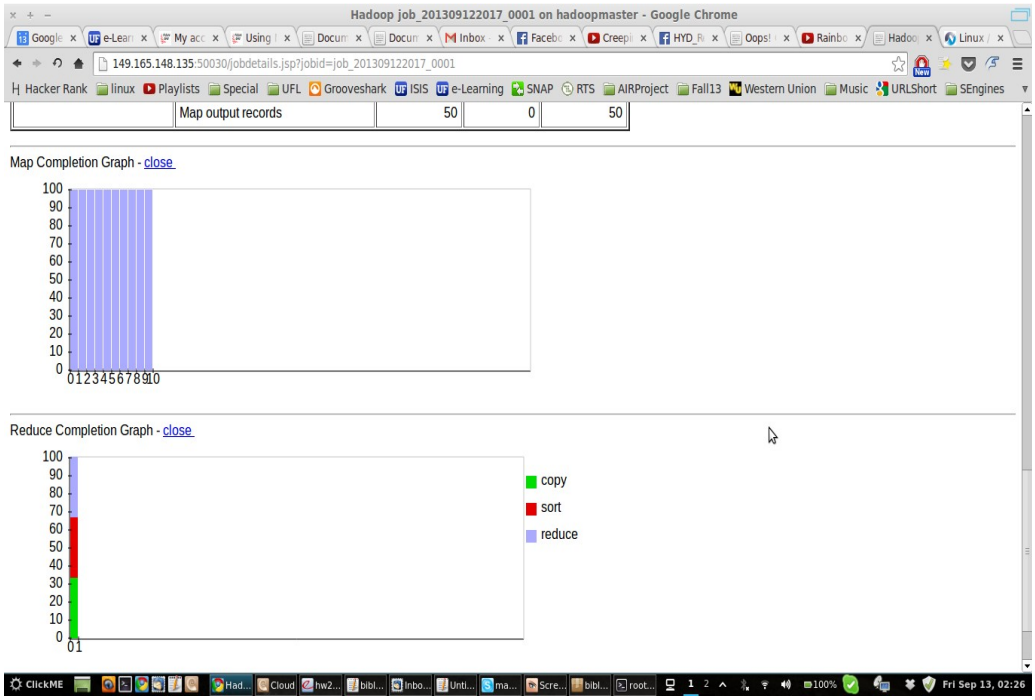
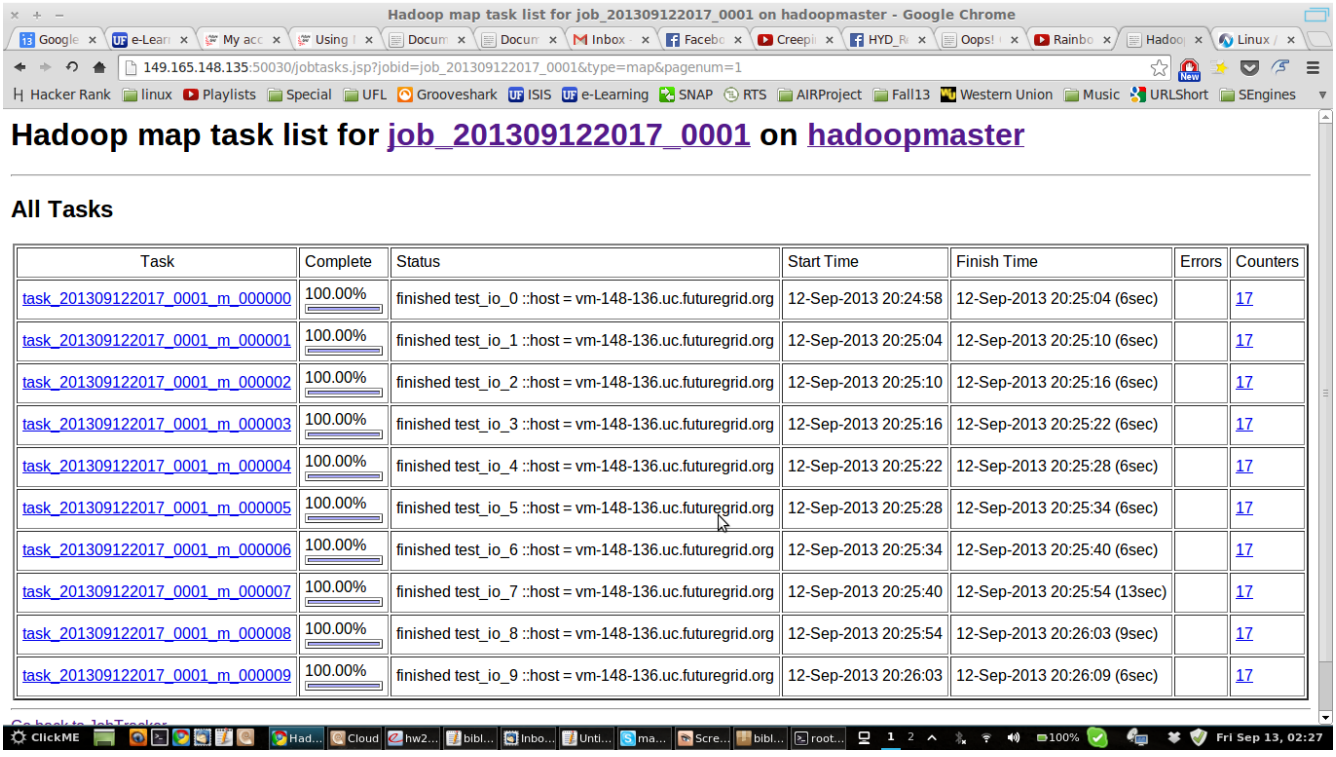


Figure-4: Map Task list



Task	Complete	Status	Start Time	Finish Time	Errors	Counters
task_201309122017_0001_m_000000	100.00%	finished test_io_0 ::host = vm-148-136.uc.futuregrid.org	12-Sep-2013 20:24:58	12-Sep-2013 20:25:04 (6sec)		17
task_201309122017_0001_m_000001	100.00%	finished test_io_1 ::host = vm-148-136.uc.futuregrid.org	12-Sep-2013 20:25:04	12-Sep-2013 20:25:10 (6sec)		17
task_201309122017_0001_m_000002	100.00%	finished test_io_2 ::host = vm-148-136.uc.futuregrid.org	12-Sep-2013 20:25:10	12-Sep-2013 20:25:16 (6sec)		17
task_201309122017_0001_m_000003	100.00%	finished test_io_3 ::host = vm-148-136.uc.futuregrid.org	12-Sep-2013 20:25:16	12-Sep-2013 20:25:22 (6sec)		17
task_201309122017_0001_m_000004	100.00%	finished test_io_4 ::host = vm-148-136.uc.futuregrid.org	12-Sep-2013 20:25:22	12-Sep-2013 20:25:28 (6sec)		17
task_201309122017_0001_m_000005	100.00%	finished test_io_5 ::host = vm-148-136.uc.futuregrid.org	12-Sep-2013 20:25:28	12-Sep-2013 20:25:34 (6sec)		17
task_201309122017_0001_m_000006	100.00%	finished test_io_6 ::host = vm-148-136.uc.futuregrid.org	12-Sep-2013 20:25:34	12-Sep-2013 20:25:40 (6sec)		17
task_201309122017_0001_m_000007	100.00%	finished test_io_7 ::host = vm-148-136.uc.futuregrid.org	12-Sep-2013 20:25:40	12-Sep-2013 20:25:54 (13sec)		17
task_201309122017_0001_m_000008	100.00%	finished test_io_8 ::host = vm-148-136.uc.futuregrid.org	12-Sep-2013 20:25:54	12-Sep-2013 20:26:03 (9sec)		17
task_201309122017_0001_m_000009	100.00%	finished test_io_9 ::host = vm-148-136.uc.futuregrid.org	12-Sep-2013 20:26:03	12-Sep-2013 20:26:09 (6sec)		17

Figure-5: Reduce Task List

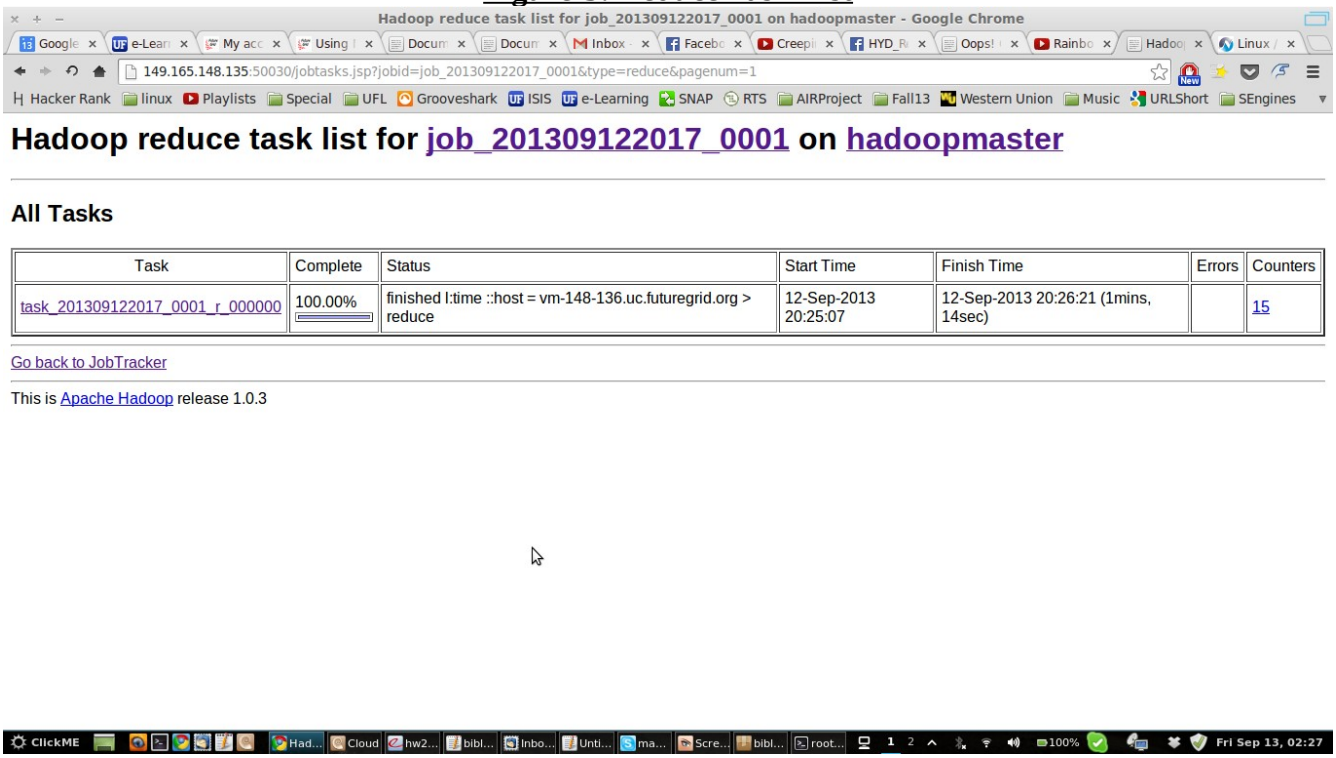
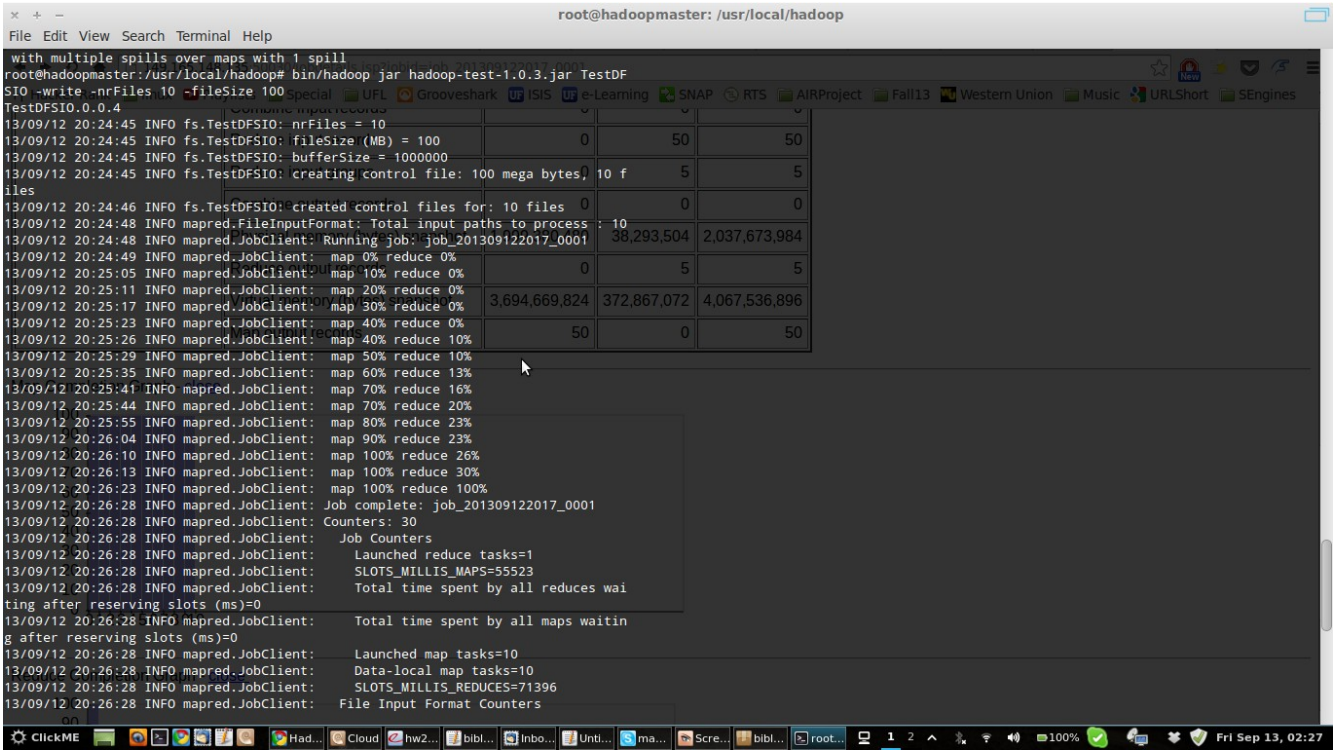


Figure-6: Screen-shot of TestDFSIO on hadoop cluster



Gaurab Dey Homework-02 UFID: 78029140
Email: gaurabde@ufl.edu

Answer-2:
Application: TestDFSIO
Details: Its benchmark testing code for HDFS read and write. It is used to detect performance bottleneck in hadoop cluster network. The output of the test can be used to optimize hardware, OS and cluster environment for better performance.

Processing: TestDFSIO is design to create only one file and one-to-one mapping from files to map tasks. Split structures are created in a way that each map service gets only one file name for write or read. It measures the three parameters for each operation:

- size of file
- time taken for the task
- rate (size/time)