**ASSIGNMENT 5**

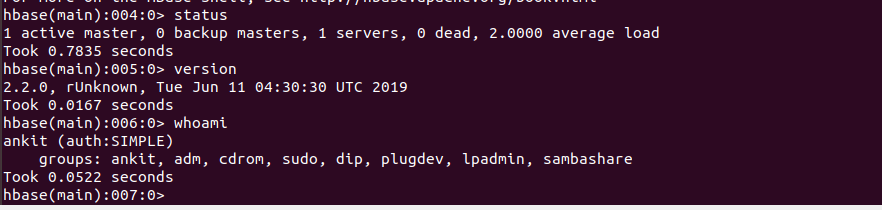
**PART 3 - Execute 2 HBase commands from each of the 6 groups** (Total 12 commands), and place the screenshots into a word file, and upload to Blackboard.

<https://learnhbase.wordpress.com/2013/03/02/hbase-shell-commands/>

**1) General  HBase shell commands**

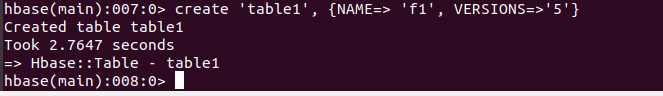
**a- hbase> status**

**b- hbase> version**

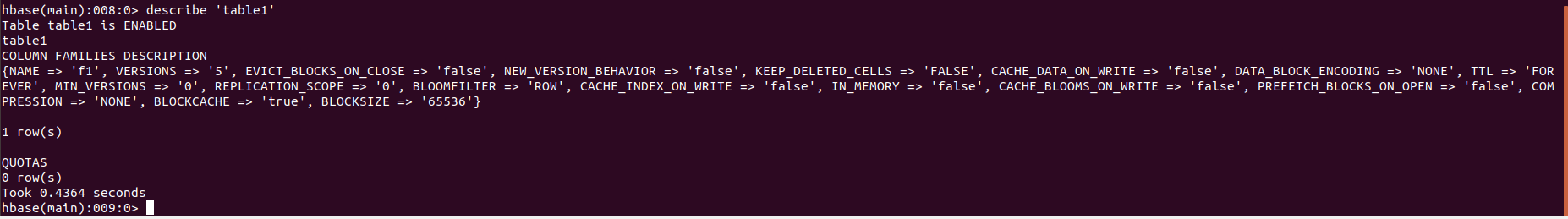


**2) Tables Management commands**

a- create

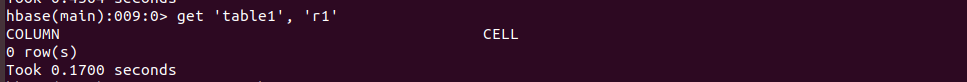


b- describe

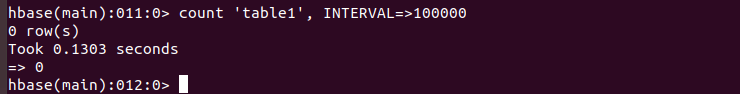


**3) Data Manipulation commands**

**a- get**

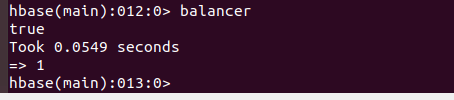


**b- count**

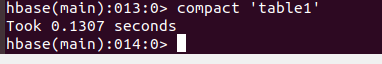


**4) HBase surgery tools**

**a- balancer**

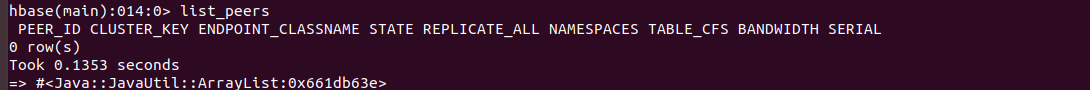


**b- compact**

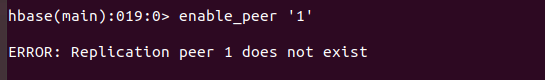


**5) Cluster replication tools**

**a- list\_peers**

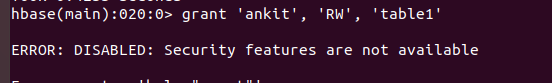


**b- enable\_peer**

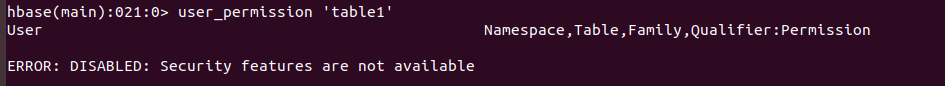


**6) Security tools**

**a- grant**



**b- user\_permission**



**PART 4 - Read the following papers (attached), and write a short summary/report.**

**Building a High-level Dataflow System on top of MapReduce: The Pig Experience**

**Summary:**

**Nova: Continuous Pig/Hadoop Workflows**

**Summary:**

**Pig Latin: A Not-So-Foreign Language for Data Processing**

**Summary:**

**PART 5 - Programming Assignment**Execute one function of your choice from each group of commands, i.e., one function from Eval Functions, one function from Load/Store functions, one function from Math functions, etc., from the Official Pig website. Every time you execute a command copy-paste the screenshot, including the output, to a word document, and submit with your assignment.  
<http://pig.apache.org/docs/r0.17.0/func.html>

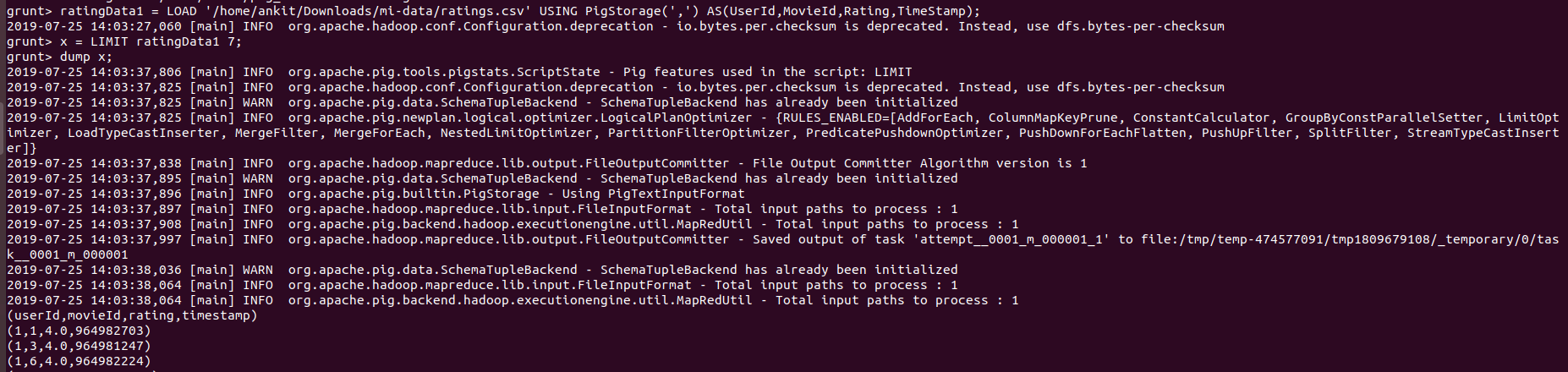
**1) EVAL Functions**

**COUNT**



**2) Load/Store Functions**

**PIGSTORAGE**

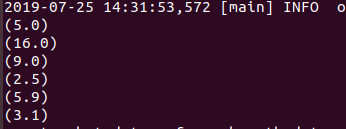


**3) Math Functions**

**CBRT**

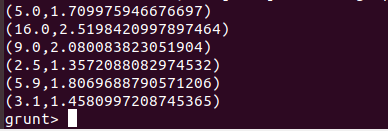












**4) String Functions**

**ENDSWITH()**

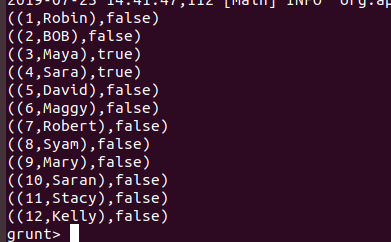












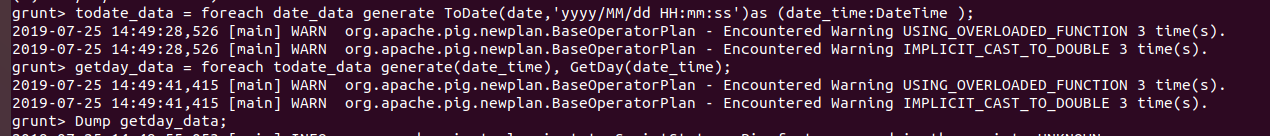
**5) Datetime Functions**

**GetDay()**





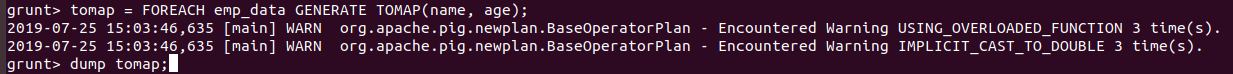


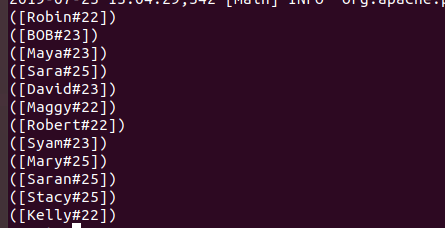




**6) Tuple,Bag,Map Functions**

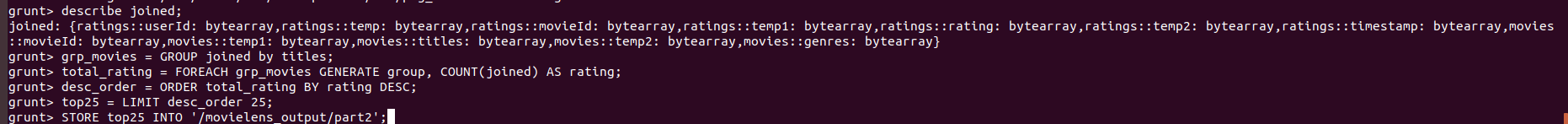
**TOMAP()**



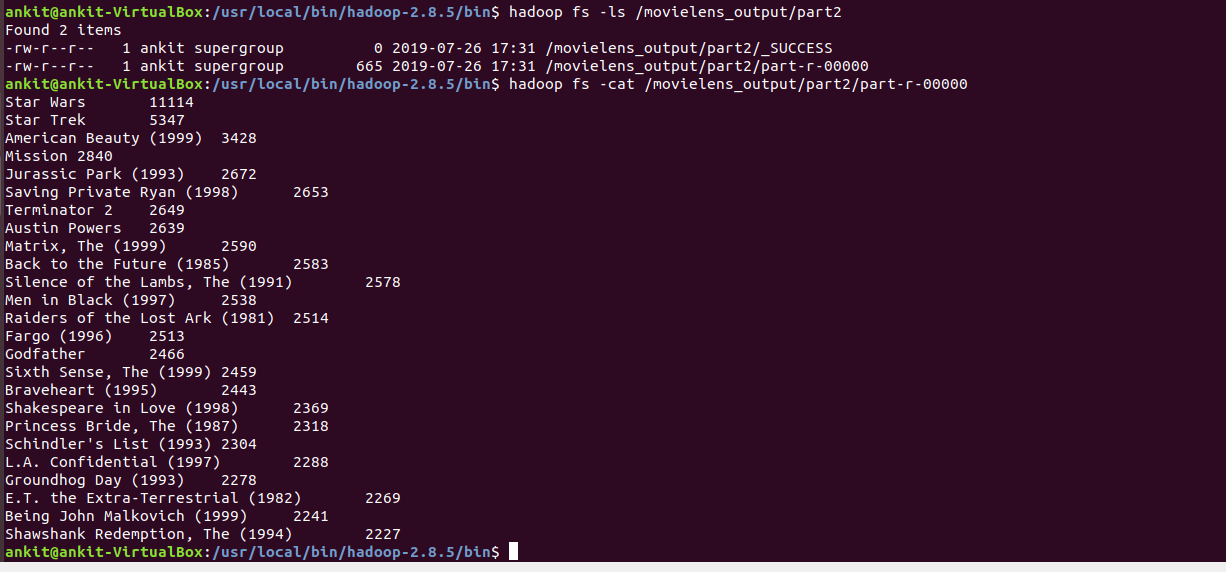


**PART 6 - Programming Assignment - Apache Pig (Use .pig scripts)**   (1 million ratings from 6000 users on 4000 movies).  
   <http://grouplens.org/datasets/movielens/>

**Task 1. Write a Pig Script to find the top 25 rated movies in the movieLens dataset**



**OUTPUT:**

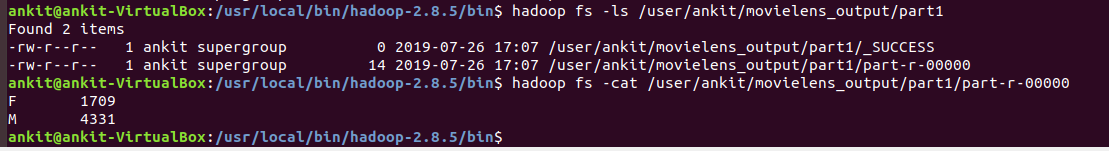


**Task 2. Write a Pig Script to find the number of males and females in the movielens dataset**





**OUTPUT:**

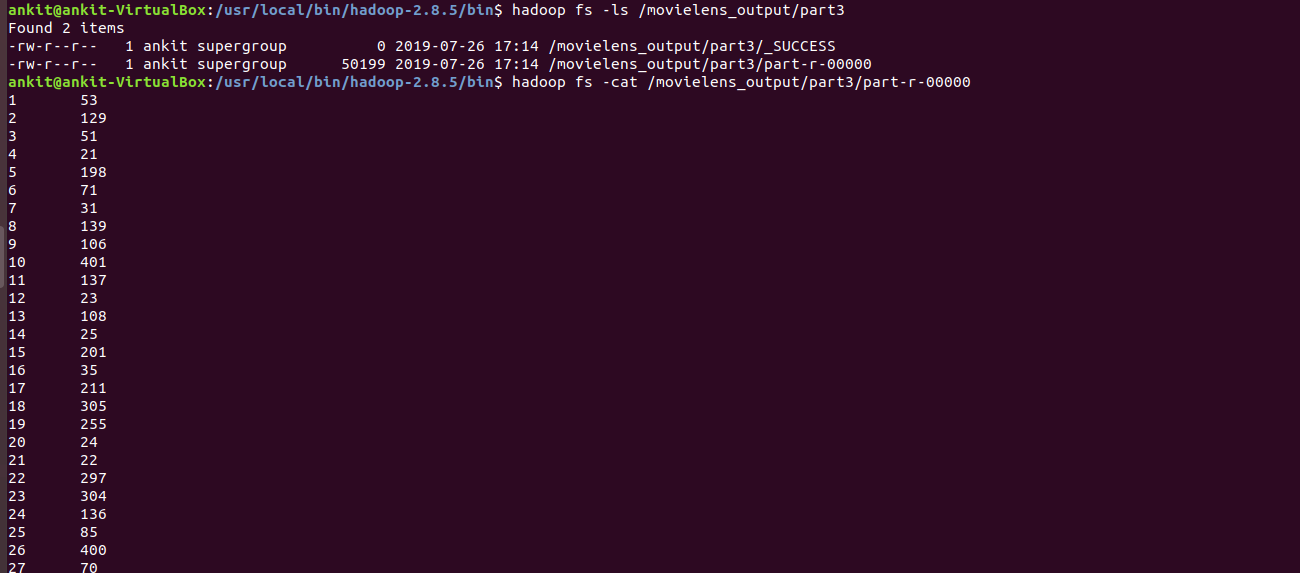


**Task 3. Write a Pig Script to find the number of movies rated by different users**





**OUTPUT:**



**PART 7 - Programming Assignment - Apache Pig (Use GRUNT Shell)**

Copy the ‘access.log’ file, used in previous assignments, into HDFS under /logs directory.  
Using the access.log file stored in HDFS, implement Pig Script to find the number of times each IP accessed the website.

**PIG Script:**

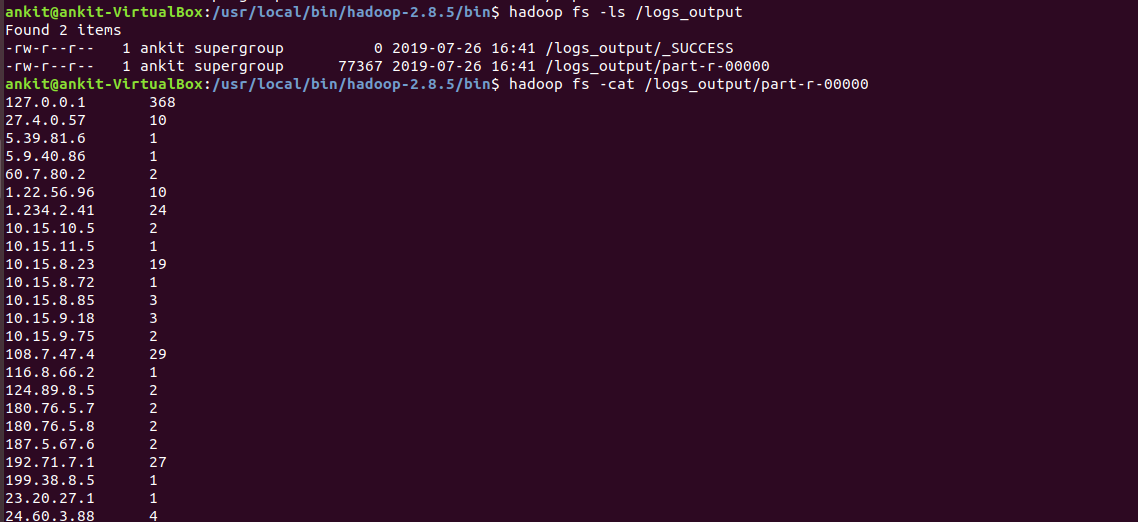








**Output:**



**ILLUSTRATE Command:**

