Attached Files:

- 📄 Analysis of HDFS Under HBase - A Facebook Messages Case Study.pdf Analysis of HDFS Under HBase - A Facebook Messages Case Study.pdf - Alternative Formats (257.776 KB)
- 📄 Paper1 - Building a HighLevel Dataflow System on top of MapReduce.pdf Paper1 - Building a HighLevel Dataflow System on top of MapReduce.pdf - Alternative Formats (888.846 KB)
- 📄 Paper2 - Nova-Continuous Pig-Hadoop Workflows.pdf Paper2 - Nova-Continuous Pig-Hadoop Workflows.pdf - Alternative Formats (1.028 MB)
- 📄 Paper3 - Pig Latin_A Not-So-Foreign Language for Data Processing.pdf Paper3 - Pig Latin_A Not-So-Foreign Language for Data Processing.pdf - Alternative Formats (619.319 KB)

PART 1 - Read the following chapters (HBase in Action)
   Ch01. Introducing HBase [https://www.safaribooksonline.com/library/view/hbase-in-action/9781617290527/kindle_split_012.html]
   Ch02. Getting started [https://www.safaribooksonline.com/library/view/hbase-in-action/9781617290527/kindle_split_013.html]
   Ch04. HBase table design [https://www.safaribooksonline.com/library/view/hbase-in-action/9781617290527/kindle_split_016.html]

PART 2 - Read the following paper (attached).
   Analysis of HDFS Under HBase: A Facebook Messages Case Study

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/

PART 4 - Read the following papers (attached), and write a short summary/report.
   Building a HighLevel Dataflow System on top of MapReduce: The Pig Experience
   Nova: Continuous Pig/Hadoop Workflows
   Pig Latin: A Not-So-Foreign Language for Data Processing

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

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
   Task 2. Write a Pig Script to find the number of males and females in the movielens dataset
   Task 3. Write a Pig Script to find the number of movies rated by different users

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.