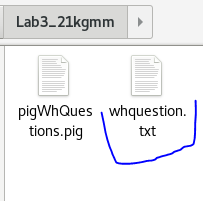
Lab\_3

Name: Karim Gamal Mahmoud Mohamed

ID: 21KGMM

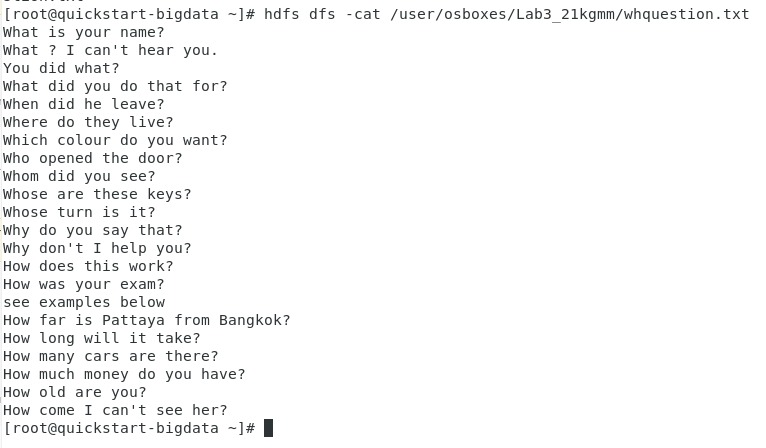
1-- Prepare a text file with WH questions by taking example sentences from the following link, name the file as 'whquestion.txt' and save it in your local drive: [WH Question Words | Vocabulary | EnglishClub.](https://www.englishclub.com/vocabulary/wh-question-words.htm)



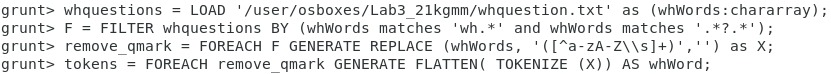
2-- Copy the whquestion.txt file into a hdfs location



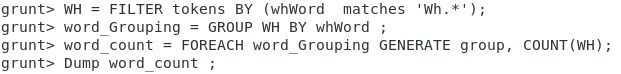




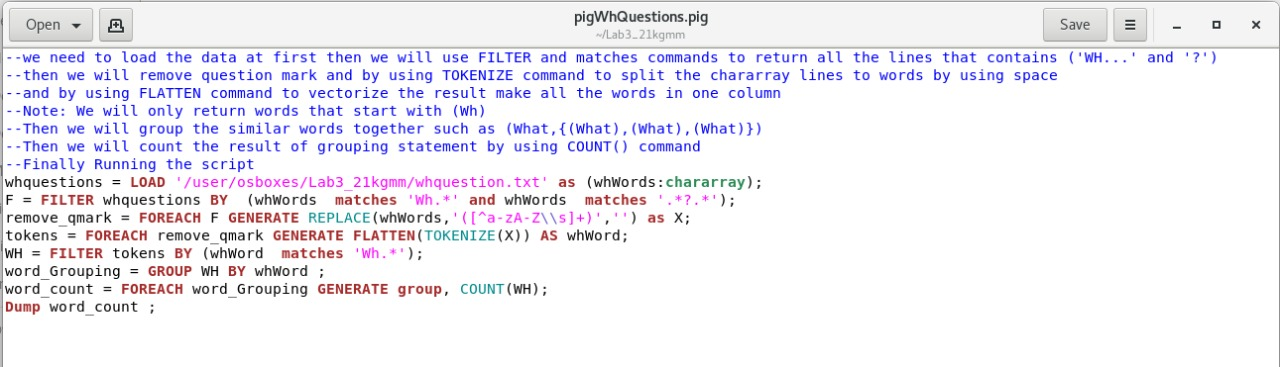
3-- Write commands in PIG to count only the frequency of the WH question words (What, Which, Who, Whom, Why...) from the question file you put in hdfs location.



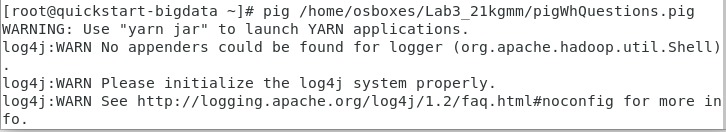


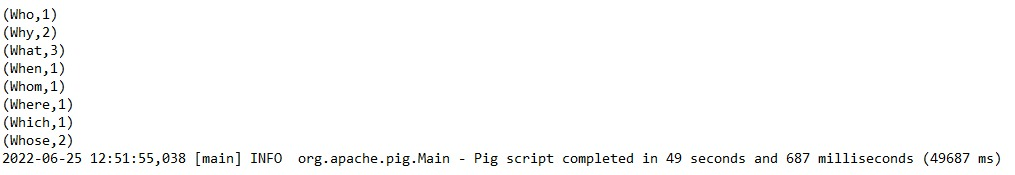


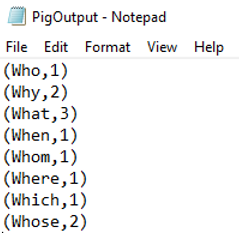
4-- Accumulate all the commands in a script file called pigwhquestion.pig



5-- Run the script file to generate the final outcome at once

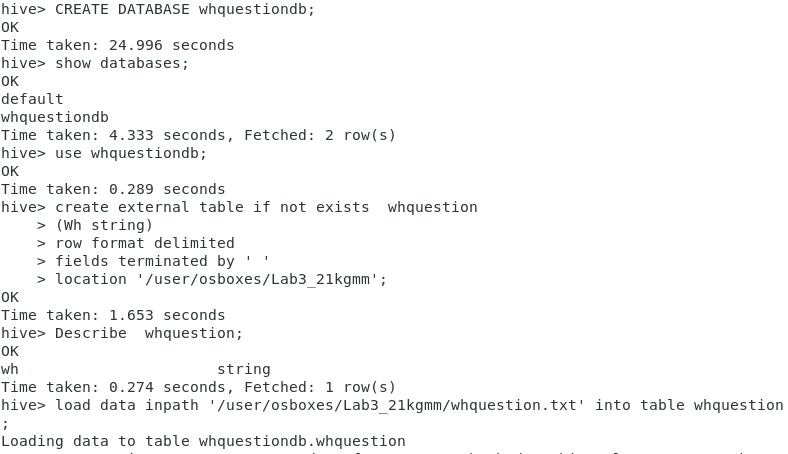


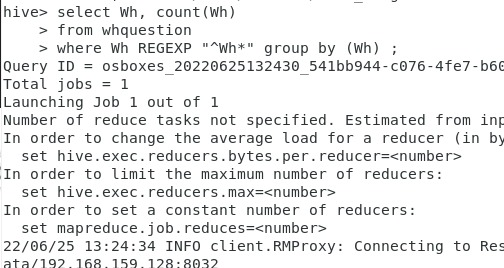


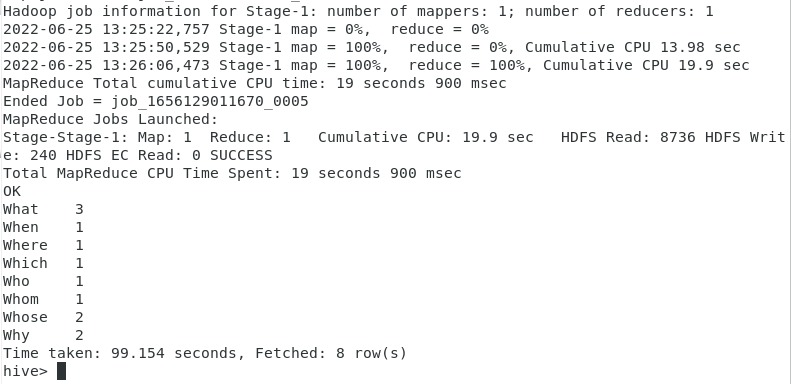


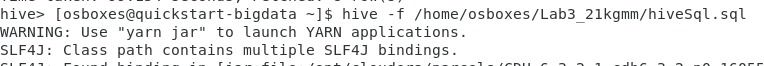
7-- Use the same input data in whquestion.txt from within Hive and give the appropriate SQL commands to solve the same word frequency count problem above. Refer to any existing material you found on the web.

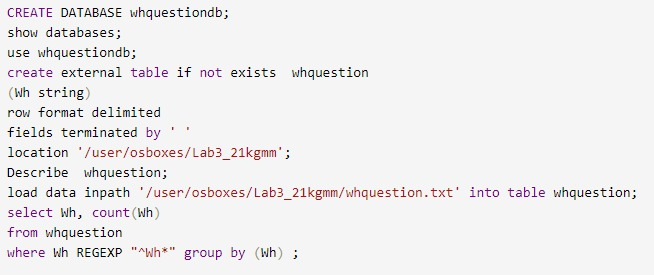
[Source:](https://riptutorial.com/hive/example/21838/word-count-example-in-hive)











Findings:

* Pig is used to perform all kinds of data manipulation operations in Hadoop. And it is abstract over MapReduce. And Pig does not support schema to store data. And not support partitioning.
* Hive is built on the top of Hadoop and is used to process structured data in Hadoop. So we should create database to store our data as input, but in pig it works on any type of data. And Hive supports schema for data insertion in tables. And supports partitioning.

Note: Hive is easier than Pig in commands, but it loads data slowly.

**Sources :**

1. [hive Tutorial => Word Count Example in Hive (riptutorial.com)](https://riptutorial.com/hive/example/21838/word-count-example-in-hive)
2. [hive Tutorial - SELECT Statement (sodocumentation.net)](https://sodocumentation.net/hive/topic/4133/select-statement)
3. [Difference between Pig and Hive - GeeksforGeeks](https://www.geeksforgeeks.org/difference-between-pig-and-hive/)
4. [Apache Pig - Running Scripts (tutorialspoint.com)](https://www.tutorialspoint.com/apache_pig/apache_pig_running_scripts.htm)