

# WQD7007 Big Data Management

## Lab Test (Hive/Pig)

Instructions: Answer all the questions. Explain adequately how you get the answers that can include: 1) codes or process used; and 2) print screens and/or related files that can justify your outcome. You can create a readme file (either in .txt or .docx) to explain your answer. Zip all the files and submit to the spectrum “Lab test” submission page.

### Part 1: (7 marks)

1. Download a dataset from the Spectrum page. Please refer to Appendix 1 (at the end of the document) on which dataset you should download. Import the downloaded dataset to HDFS.  
(3 marks)
2. By using Hive or Pig, identify 5 rows of data that have the
  - a. highest reading score.
  - b. lowest CGPA.  
(4 marks)

### Part 2: (8 marks)

1. Import text from [here](#) to HDFS.  
(1 mark)
2. Run a word count program using Hadoop MapReduce concept to count the word occurrence of the imported text as in step 1. Save the results in HDFS.  
(3 marks)
3. Import the result from step 2 to Hive or Pig. Display:
  - a. 5 words with highest counts in ascending alphabetical order.
  - b. 5 words with 5 counts in descending alphabetical order.  
(4 marks)

## Appendix 1:

Please download respective datasets (in .csv format, e.g. Set1.csv) based on the name list below:

No	Name	Set (1,2,3,4, or 5)
1.	AHMAD ZAMRIK	1
2.	AIMI NABILAH BINTI HASSIN	2
3.	ANIS 'AFIFI IZYANI BINTI AZIZAN	3
4.	CHEN,YANAN	4
5.	CHEONG SOON LOONG	5
6.	CHEY POH LIAN	1
7.	CHONG KIN HOW	2
8.	CHOONG EN JUN	3
9.	DAYANG ALMASURIN BINTI DATU ARSHAD	4
10.	FARAH NORLIYANA BINTI AZAHAR	5
11.	GOH BIAN CHIAT	1
12.	HII LEE ZING	2
13.	KUAN NAM SUNG	3
14.	KUMARMANGAL ROY	4
15.	LEE KWAN LI	5
16.	LEE MEI TEE	1
17.	LEOW WEI QIN	2
18.	LIM JOU HUI	3
19.	LIM YAN	4
20.	LIU,HONGYANG	5
21.	MAS RAHAYU BINTI MOHAMAD	1
22.	MOHAMED FATHI MOHAMED HAMAD	2
23.	MOHD FIRDAUS BIN SHAHRIN	3
24.	MUHAMMAD AMIRUL SAMSUL BAHAR	4
25.	MUHAMMAD FAIZ BIN ABDULLAH	5
26.	MUHAMMAD HAFIZ BIN KHAIRUDIN	1
27.	NADIA BELLA GUSTIANI RASYIDI	2
28.	NG JIUNN JYE	3
29.	NIK SYEN CHYN	4
30.	NUR 'AISYAH BINTI ZAKARIA ADLI	5
31.	NUR AZILAH BAHARUDDIN	1
32.	NUR HUMAIRA BINTI SULAIMAN	2
33.	NURBAYANI BINTI HALIM	3
34.	NURFARHANA BINTI OMAR	4
35.	NURSHARA FAZREEN MAZLAN	5
36.	NURUL SYAFIQAH BINTI MD KHAIRI	1
37.	NURZUHRAH BINTI HASSAN	2
38.	NYANARAAM KANAISAN	3
39.	OOI LI JUAN	4
40.	PARVITRA A/P SOMAHSUNDRAM	5
41.	RAJA NADIA BINTI RAJA MUSAFARUDIN	1

42.	RAJIB KANTI PAL	2
43.	RAYMOND OOI KHAI BOON	3
44.	RINASHINI A/P ARUNASALAM SUKORMARU	4
45.	RISHI NIGAM	5
46.	SENG JIA XIN	1
47.	SITI ZULAIKHA BINTI DOLLAH	2
48.	SOFIA ATIKAH BINTI MOHD ZABIT	3
49.	SRITHARAN SIVAGURU	4
50.	STEVEN CHONG SHEN YEN	5
51.	TAN CHANG JUNG	1
52.	TAN JOR YI	2
53.	TAN QIAN HUI	3
54.	TAN SIA HONG	4
55.	TARASHINI A/P SUTHESAN	5
56.	WAI CHEE SAN	1
57.	WONG WAI CHENG	2
58.	YAP BOON HUI	3
59.	ZAIMIE AZMIN BIN ZAINUL ABIDIN	4
60.	ZARRIEF SOFEA BINTI ABDUL WAHID	5
61.	ZULKANAIN BIN HASAN	1