

CSE3BDC Marking sheet 2015

Student name:	<u>Junqi Li</u>	Days late	<u>0</u>
Student id:	<u></u>	Total Mark	<u>106</u>
	<u></u>	(- deductions)	<u></u>
		Final Mark	<u>100</u>

Task 01	
A) Correct output (6 marks)	6
A) Correct combiner/local aggregation (3 marks)	3
B) Correct output (6 marks)	6
B) Too many tables (-1 marks)	
C) Correct output (4 marks)	4
C) Handle 0 % (2 marks)	0
C) Too many tables (- 1 marks)	
D) Correct output (7 marks)	7
D) Everything done in reducer (-7 marks)	
D) Correct combiner/local aggregation (3 marks)	3
E) Correct output (10 marks)	10
E) Check code is reasonable	
<i>Total ([/])</i>	39

Task 02	
A) Correct output (5 marks)	5
A) Using reduce with max function or top (1 mark) - using sorting is less efficient than reduce	1
A) Poor coding style (-3 marks)	
B Spark) Correct output (4 marks)	4
B Spark) Used reduceByKey instead of groupBy (1 marks)	1
B Spark) Using reduce with max function or top (1 mark) - using sorting is less efficient than reduce	0
B Spark) Poor coding style (-3 marks)	
B Hive) Correct output (6 marks)	6

CSE3BDC Marking sheet 2015

C) Selection in Mapper (3 marks)	3
C) Output correct (9 marks)	9
C) Used combiner/local aggregation (4 marks)	4
<i>Total ([/i])</i>	33

Task 03	
A) Correct output (6 marks)	6
A) Poor coding style (-3 marks)	
B) Correct grouping for small file (2 marks)	2
B) Correct sorting based on word for small file (2 marks)	2
B) Correct sorting based on count for small file (2 marks)	2
B) Correct output for large file (3 marks)	
B) Poor coding style (-3 marks)	3
C) Load object file correctly (2 marks)	2
C) Cached data (1 marks)	0
C) Correct output (2 marks)	2
<i>Total ([/i])</i>	19

Misc	
Bonus 1) Correct output for small data set (6 marks)	6
Bonus 1) Large data file finishes within 5 minutes without running out of memory (4 marks)	4
Bonus 2)	
Bonus 2)	
Bonus 2)	
Bonus 2)	
Taking advantage of MapReduce Framework sorting in at least one question (5 marks)	5
<i>Total ([/i])</i>	15