Homework1-data mining

2016年3月22日　星期二　　　下午11:39

Assignment:
Matrix multiplication:
A (10000*10000) * B(10000*10000)

Requirements:
1. Using spark (Pyspark, scala, Java) to design the map-reduce method of multiplication
2. Write the result as the key value pair ((row_number, col_number), value)
3. Give me the top 100 pair result (part of row 1)
4. Assign the code in the report
5. Extra bonus: optimization with splitting the matrix into few pieces and do the whole map-reduce multiplication
6. Name your report like this studentid_hw_number.pdf

External links and coding reference:
CMU assignment slides (can skip to the example)
https://docs.google.com/presentation/d/1Xn1rYdX7LKoO06mQJYKqM2QNSFo_p5r86xyHmekgxmY/mobilepresent?slide=id.g10bb0037c0_0_53

Books:
Learning spark
http://download.csdn.net/detail/szlhj/8444411