Program instruction

yupeng.gao

February 2016

1 Before running

Put the mutualFriends.jar in your campus server disk.

2 Question1

Input file path: $/yxg140730/input/soc - LiveJournal1Adj.txt$
output path: $/yxg140730\_out$

enter the path of where the Jar file located. code like this in the commend line in your cs6360 server:

$hadoop\ dfs\ -rm\ -r\ /yxg140730\_out$
$hadoop\ jar\ mutualFriends.jar\ mutualFriend$
$hadoop\ dfs\ -cat\ /yxg140730\_out/*$

Then you should be able to see the result.

3 Question2

input file path: $/yxg140730/input/resultFromQusOne$
output file path:$/yxg140730\_out$

Both of the path have been encoded. So we only need to provide two use ID. type in the commend line:

$hadoop\ dfs\ -rm\ -r\ /yxg140730\_out$
$hadoop\ jar\ mutualFriends.jar\ Question2\ 1\ 35621$
$hadoop\ dfs\ -cat\ /yxg140730\_out/*$

4 Question3

job1 input file path: $/yxg140730/input/resultFromQusOne$
  job1 output file path: $/yxg140730/intermediate_output$
  job2 input file path: $/yxg140730/input1/userdata.txt$
  job2 output file path: $/yxg140730\_out$

The path have been encoded in program. Type in the commend line:

$hadoop\ dfs\ -rm\ -r\ /yxg140730\_out$
$hadoop\ dfs\ -rm\ -r\ /yxg140730/intermediate\_output$
$hadoop\ jar\ mutualFriends.jar\ Question3\ 1\ 35621$
$hadoop\ dfs\ -cat\ /yxg140730\_out/*$

5 Question4

job1 input file path: $/yxg140730/input/soc-LiveJournal1Adj.txt$
  job1 in-memory join file path: $/yxg140730/input1/userdata.txt$
  job1 output file path: $/yxg140730/intermediate$
  job2 input file1 path: $/yxg140730/intermediate/part-r-00000$
  job2 input file2 path:$/yxg140730/input1/userdata.txt$
  job2 output file path: $/yxg140730\_out$

type in the commend line:

$hadoop\ dfs\ -rm\ -r\ /yxg140730\_out$
$hadoop\ dfs\ -rm\ -r\ /yxg140730/intermediate/$
$hadoop\ jar\ mutualFriends.jar\ Question4$
$hadoop\ dfs\ -cat\ /yxg140730\_out/*$