Homework5

Ming Xia

November 2021

1   Introduction

In this project, I install and set up the Hadoop in ubuntu OS.

it is easy to mapper the data, and I design the output data in:

it(disambiguation)visit 0

$it_($disambiguation$)size 7471$

$Category : Merchants visit 3$

$Category : Merchants size 7286$

$Category : Merchants visit 3$

$Category : Merchants size 7286$

$Category : Merchants visit 2$

$Category : Merchants size 7286$

$Category : Merchants visit 1$

$Category : Merchants size 7286$

$Category : Merchants visit 6$

$Category : Merchants size 7286$

and then use the reducer.py to add them together.