

PS27 LinXiLi

10.5

The Problem of this code is collect() works on single node.
then execute map, reduce functions. which is not ~~the~~ Parallel.

10.13.

import sys

for line in sys.stdin

key, value = line.split(':', 1)

for word in value.strip().split()

Print '{s}\t{t}': format(word, 1)

reduce

total = 0

for line in sys.stdin

k, v = line.split(':', 1)

first blank is flagMap, (count = int(v.strip)) if key == k
total += count

third one is tuple <> (word, 1), else

Print '{s}\t{t}': format

(key, total)

key = k

total += count

if key

Print '{s}\t{t}': format(key, total)

