

RDD B

RDD C

 $rdd.map((a,b) \Rightarrow (a, b+1))$

rdd.reduceByKey(_ + _)

Partition 1

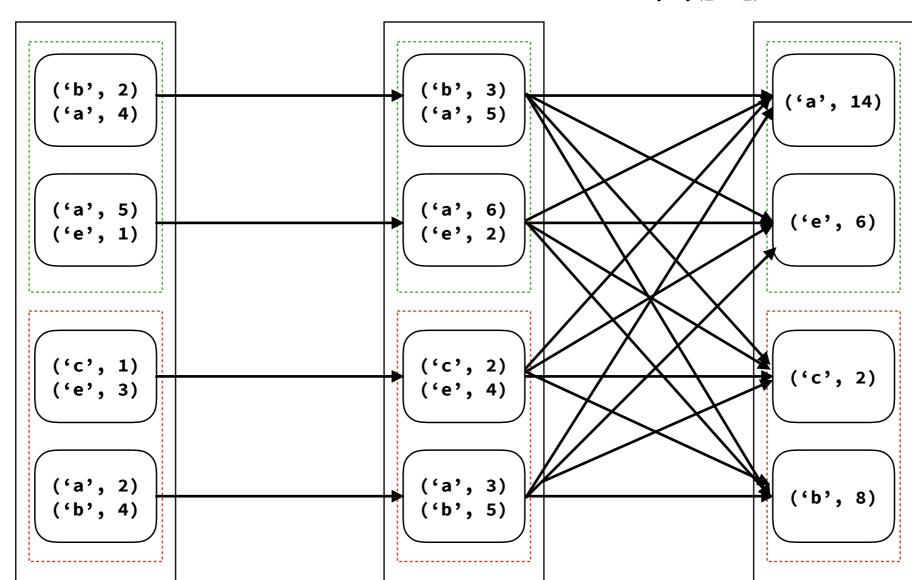
Machine 1

Partition 2

Partition 3

Machine 2

Partition 4



Narrow Dependency

Wide Dependency