R1

R0

Part-1

Part-1

Part-1

Part-0

Part-0

Part-0

Part-1

Part-0

Map

Map

Map

FlatMap

map

map

FlatMap

FlatMap

map

Map

FlatMap

map

10 GM RAM, 4 cores: 40 partitions

Driver: InputFormatClass: how many inputSplit: 8K >> 8K partitions(LogRDD)

NameNode: metadata of HDFS

10 GM RAM, 4 cores, 40 partitions

10 GM RAM, 4 cores, 40 partitions

10 GM RAM, 4 cores, 40 partitions