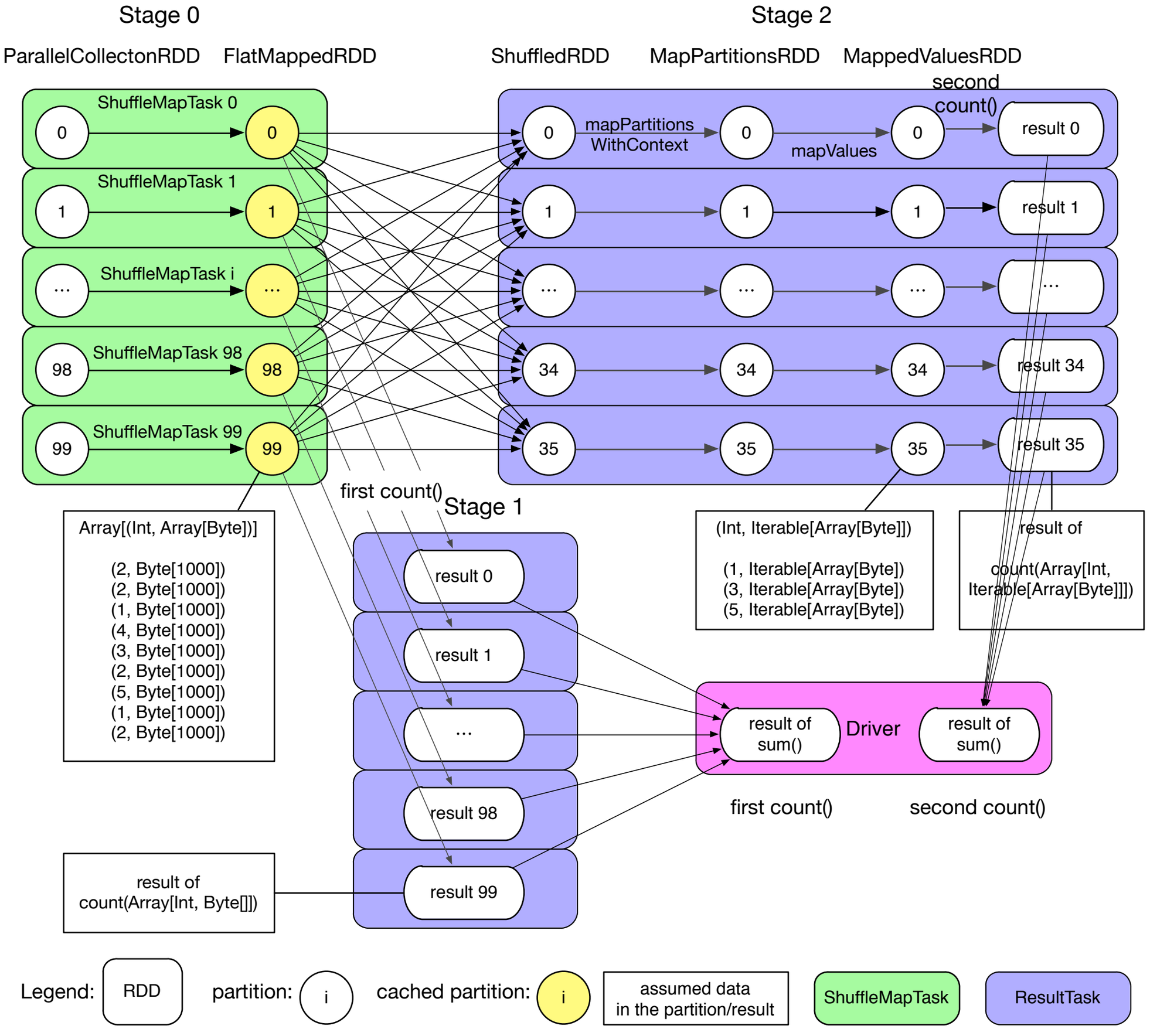
Stage 0
Stage 2
ParallelCollectonRDD
FlatMappedRDD
ShuffledRDD
MapPartitionsRDD
MappedValuesRDD
second count()
ShuffleMapTask 0
ShuffleMapTask 1
ShuffleMapTask i
ShuffleMapTask 98
ShuffleMapTask 99
mapPartitions WithContext
mapValues
result 0
result 1
result 34
result 35
first count()
Stage 1
Array[(Int, Array[Byte])]
(2, Byte[1000])
(2, Byte[1000])
(1, Byte[1000])
(4, Byte[1000])
(3, Byte[1000])
(2, Byte[1000])
(5, Byte[1000])
(1, Byte[1000])
(2, Byte[1000])
(Int, Iterable[Array[Byte]])
(1, Iterable[Array[Byte])
(3, Iterable[Array[Byte])
(5, Iterable[Array[Byte])
result of count(Array[Int, Iterable[Array[Byte]]])
result of sum()
Driver
result of sum()
first count()
second count()
result of count(Array[Int, Byte[]])
Legend:
RDD
partition:
cached partition:
assumed data in the partition/result
ShuffleMapTask
ResultTask