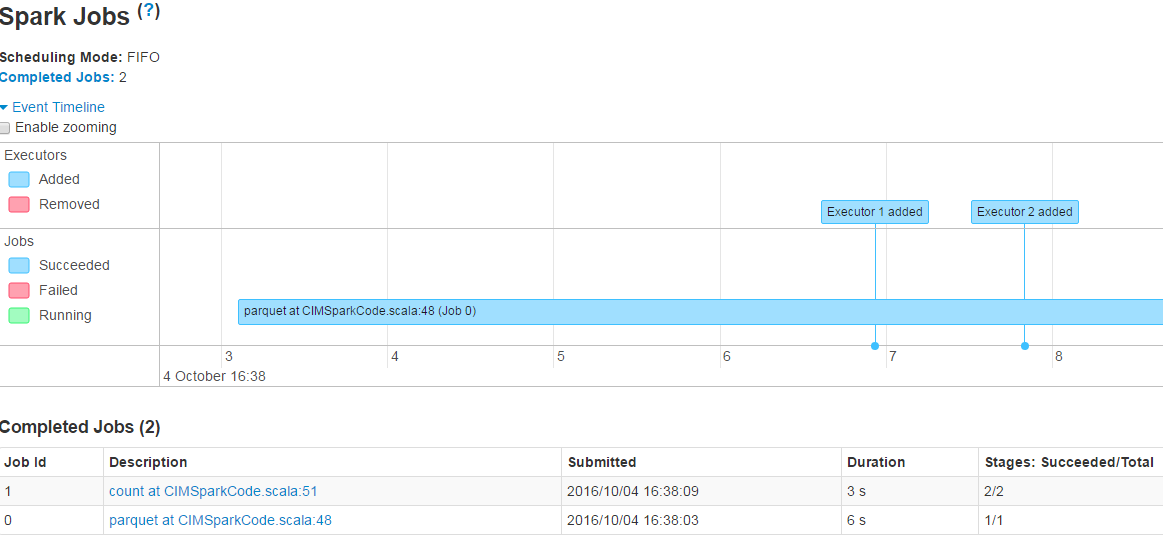
Spark Learning:

When Spark Job initialized following operations will happen:

1. Creates no.of Executors
2. No.Of Jobs
   1. 1st Job: Read data from table
   2. 2nd Job: Count the data



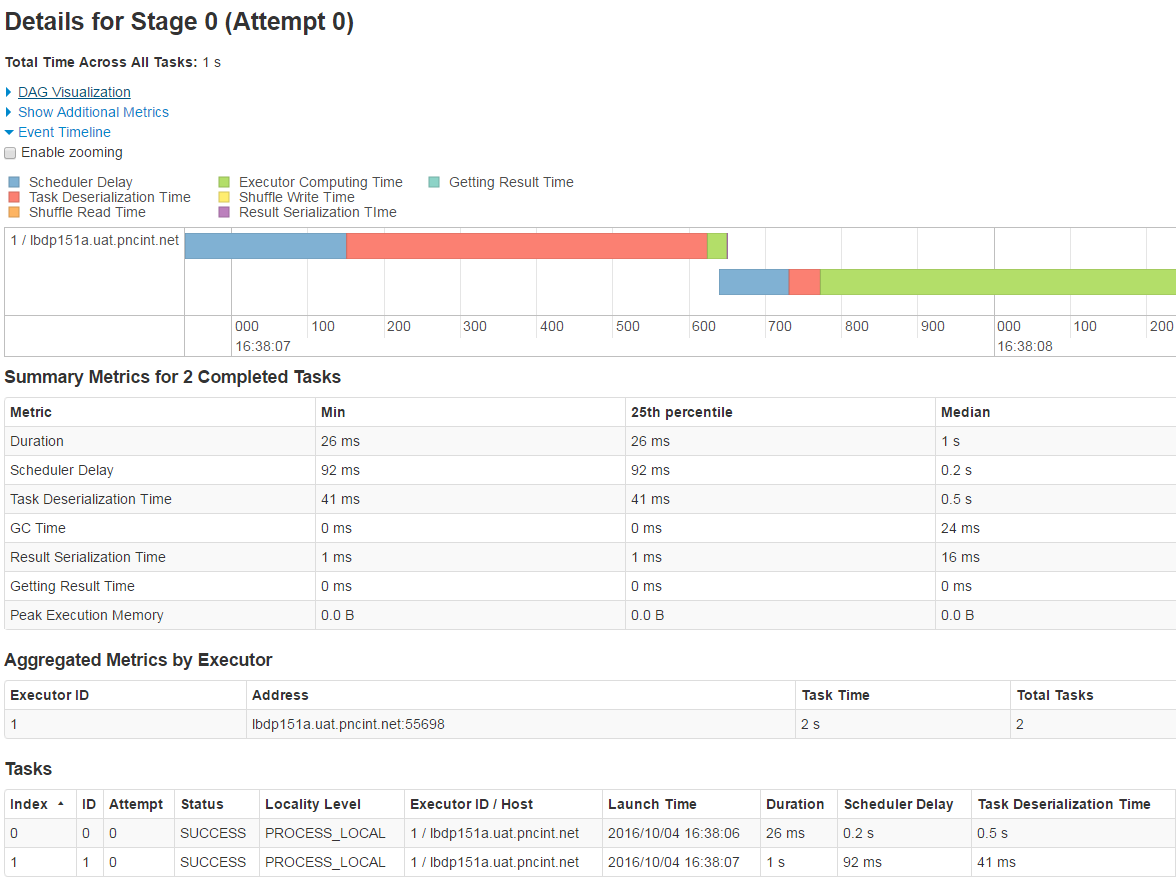
1st Job:

For each job, it creates Stages. Only 1 Stage is created.



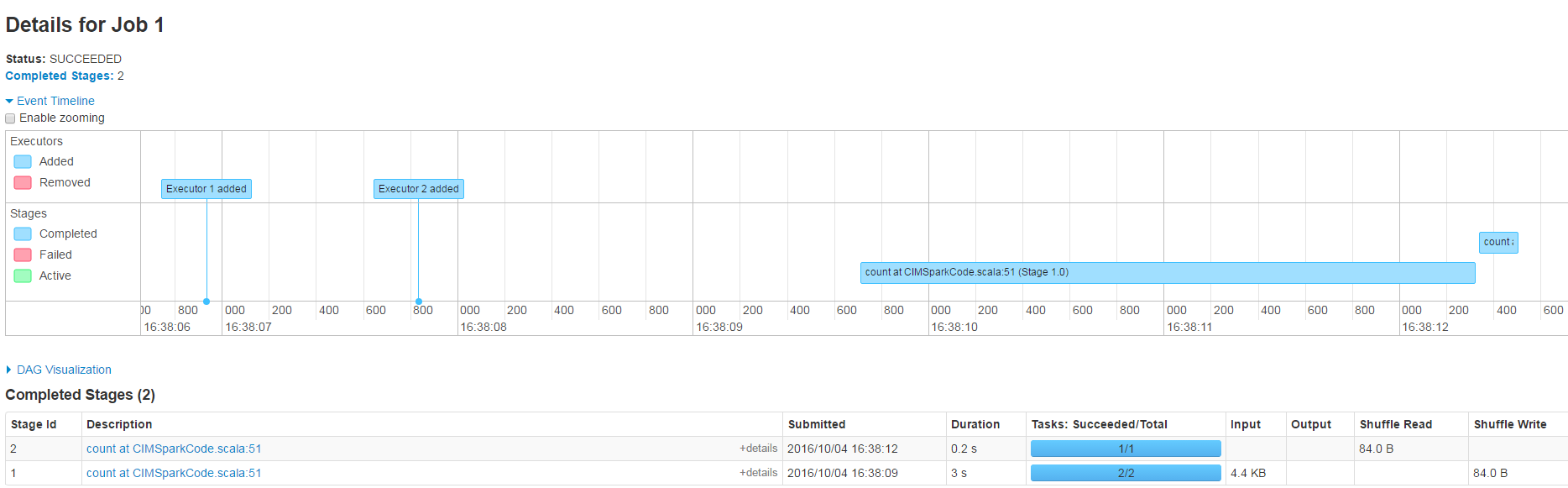
Stage:

For each stage, it creates Executor’s. Only 1 Executor Created.



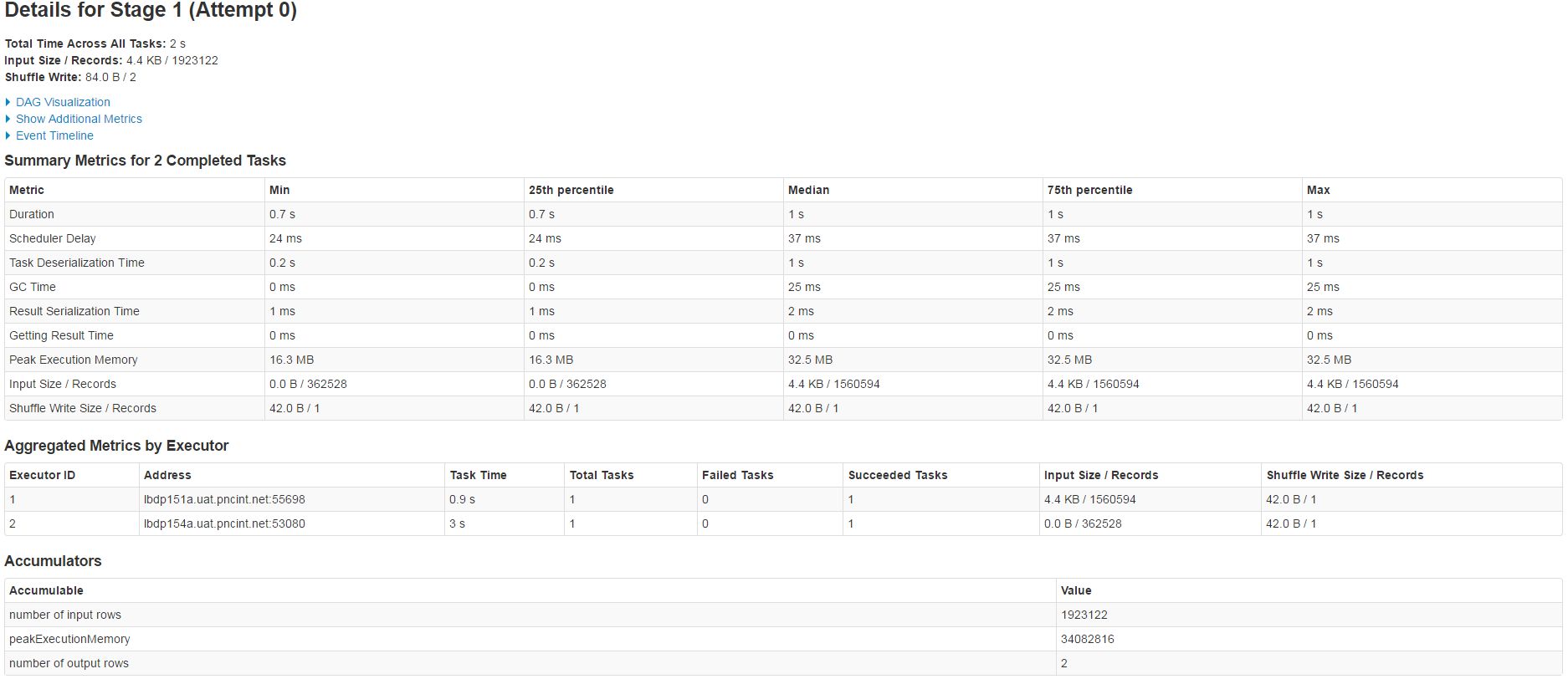
2nd Job:

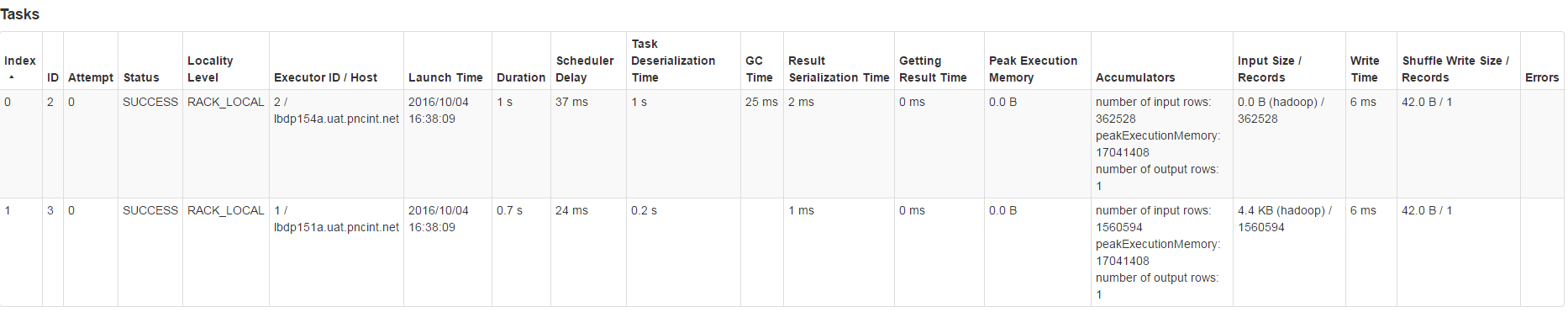
It created 2 Stages for Count operation.



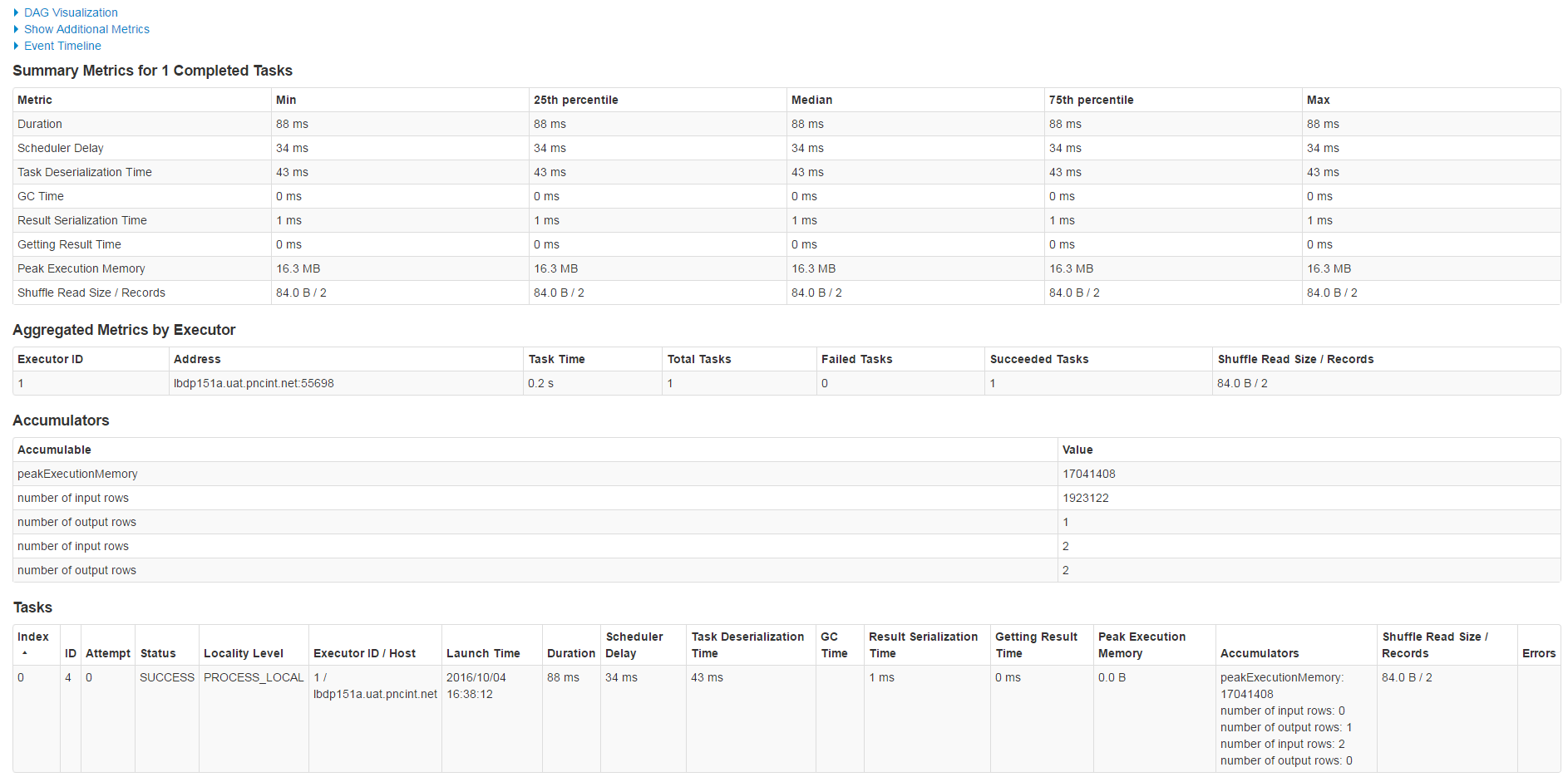
Stage 1:

There are 2 Executor’s are created. Each executor read data from Each Data Node.



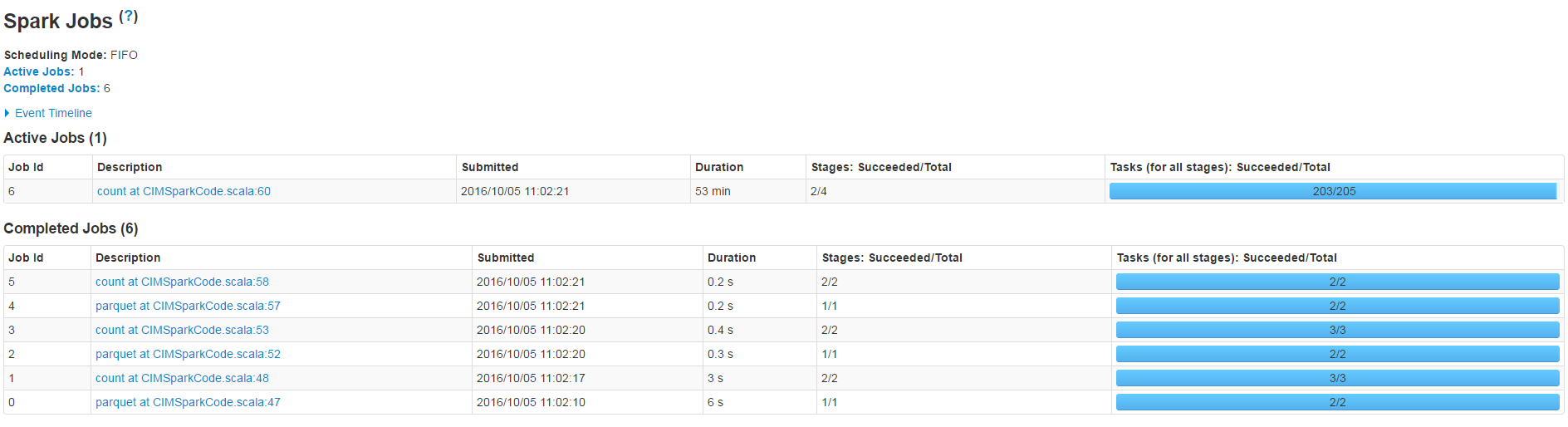


2nd Executor: Read Data from 1st Executor’s



The following Spark Job reading data from 3 tables. Each table read data and count. And join 2 tables & count.

7 jobs are created.



== Parsed Logical Plan ==

Aggregate [count(1) AS count#120L]

Join LeftOuter, Some((DATE\_KEY#39 = DATE\_KEY#99))

Relation[atm\_hier\_key#0,channel\_key#1,cust\_prospect\_type\_key#2,offer\_key#3,location\_key#4,agent\_key#5,nbr\_responses#6,nbr\_accept\_responses#7,nbr\_decline\_responses#8,nbr\_maybe\_later\_responses#9,nbr\_offers\_presented#10,nbr\_successful\_responses#11,nbr\_qualified\_offers#12,nbr\_qualified\_offers\_top\_3#13,nbr\_qualified\_offers\_nba#14,nbr\_qualified\_offers\_pos\_2#15,nbr\_qualified\_offers\_pos\_3#16,nbr\_nba\_presented#17,cif\_key#18,cif\_perm\_key#19,session\_id#20,offer\_qualification\_rank#21,offer\_display\_rank#22,nbr\_control\_group\_offers#23,nbr\_no\_offer\_offers#24,nbr\_expedite\_offers#25,card\_number#26,nbr\_qualified\_offers\_notofferd#27,sub\_disposition\_key#28,call\_key#29,branch\_id#30,employee\_id#31,bank#32,disposition\_key#33,device\_key#34,business\_segment\_key#35,operating\_system\_key#36,created\_date#37,hdfs\_load\_ts#38,date\_key#39] ParquetRelation[hdfs://bdp-isilonuat01.uat.pncint.net:8020/bdpd/bdh/01/str/pub/bdhhd01d\_bdh/cust\_intractn\_event\_fact/date\_key=20160822]

Relation[atm\_hier\_key#60,channel\_key#61,cust\_prospect\_type\_key#62,offer\_key#63,location\_key#64,agent\_key#65,nbr\_responses#66,nbr\_accept\_responses#67,nbr\_decline\_responses#68,nbr\_maybe\_later\_responses#69,nbr\_offers\_presented#70,nbr\_successful\_responses#71,nbr\_qualified\_offers#72,nbr\_qualified\_offers\_top\_3#73,nbr\_qualified\_offers\_nba#74,nbr\_qualified\_offers\_pos\_2#75,nbr\_qualified\_offers\_pos\_3#76,nbr\_nba\_presented#77,cif\_key#78,cif\_perm\_key#79,session\_id#80,offer\_qualification\_rank#81,offer\_display\_rank#82,nbr\_control\_group\_offers#83,nbr\_no\_offer\_offers#84,nbr\_expedite\_offers#85,card\_number#86,nbr\_qualified\_offers\_notofferd#87,sub\_disposition\_key#88,call\_key#89,branch\_id#90,employee\_id#91,bank#92,disposition\_key#93,device\_key#94,business\_segment\_key#95,operating\_system\_key#96,created\_date#97,hdfs\_load\_ts#98,date\_key#99] ParquetRelation[hdfs://bdp-isilonuat01.uat.pncint.net:8020/bdpd/bdh/01/str/pub/bdhhd01d\_bdh/cust\_intractn\_event\_fact]

== Analyzed Logical Plan ==

count: bigint

Aggregate [count(1) AS count#120L]

Join LeftOuter, Some((DATE\_KEY#39 = DATE\_KEY#99))

Relation[atm\_hier\_key#0,channel\_key#1,cust\_prospect\_type\_key#2,offer\_key#3,location\_key#4,agent\_key#5,nbr\_responses#6,nbr\_accept\_responses#7,nbr\_decline\_responses#8,nbr\_maybe\_later\_responses#9,nbr\_offers\_presented#10,nbr\_successful\_responses#11,nbr\_qualified\_offers#12,nbr\_qualified\_offers\_top\_3#13,nbr\_qualified\_offers\_nba#14,nbr\_qualified\_offers\_pos\_2#15,nbr\_qualified\_offers\_pos\_3#16,nbr\_nba\_presented#17,cif\_key#18,cif\_perm\_key#19,session\_id#20,offer\_qualification\_rank#21,offer\_display\_rank#22,nbr\_control\_group\_offers#23,nbr\_no\_offer\_offers#24,nbr\_expedite\_offers#25,card\_number#26,nbr\_qualified\_offers\_notofferd#27,sub\_disposition\_key#28,call\_key#29,branch\_id#30,employee\_id#31,bank#32,disposition\_key#33,device\_key#34,business\_segment\_key#35,operating\_system\_key#36,created\_date#37,hdfs\_load\_ts#38,date\_key#39] ParquetRelation[hdfs://bdp-isilonuat01.uat.pncint.net:8020/bdpd/bdh/01/str/pub/bdhhd01d\_bdh/cust\_intractn\_event\_fact/date\_key=20160822]

Relation[atm\_hier\_key#60,channel\_key#61,cust\_prospect\_type\_key#62,offer\_key#63,location\_key#64,agent\_key#65,nbr\_responses#66,nbr\_accept\_responses#67,nbr\_decline\_responses#68,nbr\_maybe\_later\_responses#69,nbr\_offers\_presented#70,nbr\_successful\_responses#71,nbr\_qualified\_offers#72,nbr\_qualified\_offers\_top\_3#73,nbr\_qualified\_offers\_nba#74,nbr\_qualified\_offers\_pos\_2#75,nbr\_qualified\_offers\_pos\_3#76,nbr\_nba\_presented#77,cif\_key#78,cif\_perm\_key#79,session\_id#80,offer\_qualification\_rank#81,offer\_display\_rank#82,nbr\_control\_group\_offers#83,nbr\_no\_offer\_offers#84,nbr\_expedite\_offers#85,card\_number#86,nbr\_qualified\_offers\_notofferd#87,sub\_disposition\_key#88,call\_key#89,branch\_id#90,employee\_id#91,bank#92,disposition\_key#93,device\_key#94,business\_segment\_key#95,operating\_system\_key#96,created\_date#97,hdfs\_load\_ts#98,date\_key#99] ParquetRelation[hdfs://bdp-isilonuat01.uat.pncint.net:8020/bdpd/bdh/01/str/pub/bdhhd01d\_bdh/cust\_intractn\_event\_fact]

== Optimized Logical Plan ==

Aggregate [count(1) AS count#120L]

Project

Join LeftOuter, Some((DATE\_KEY#39 = DATE\_KEY#99))

Project [DATE\_KEY#39]

Relation[atm\_hier\_key#0,channel\_key#1,cust\_prospect\_type\_key#2,offer\_key#3,location\_key#4,agent\_key#5,nbr\_responses#6,nbr\_accept\_responses#7,nbr\_decline\_responses#8,nbr\_maybe\_later\_responses#9,nbr\_offers\_presented#10,nbr\_successful\_responses#11,nbr\_qualified\_offers#12,nbr\_qualified\_offers\_top\_3#13,nbr\_qualified\_offers\_nba#14,nbr\_qualified\_offers\_pos\_2#15,nbr\_qualified\_offers\_pos\_3#16,nbr\_nba\_presented#17,cif\_key#18,cif\_perm\_key#19,session\_id#20,offer\_qualification\_rank#21,offer\_display\_rank#22,nbr\_control\_group\_offers#23,nbr\_no\_offer\_offers#24,nbr\_expedite\_offers#25,card\_number#26,nbr\_qualified\_offers\_notofferd#27,sub\_disposition\_key#28,call\_key#29,branch\_id#30,employee\_id#31,bank#32,disposition\_key#33,device\_key#34,business\_segment\_key#35,operating\_system\_key#36,created\_date#37,hdfs\_load\_ts#38,date\_key#39] ParquetRelation[hdfs://bdp-isilonuat01.uat.pncint.net:8020/bdpd/bdh/01/str/pub/bdhhd01d\_bdh/cust\_intractn\_event\_fact/date\_key=20160822]

Project [DATE\_KEY#99]

Relation[atm\_hier\_key#60,channel\_key#61,cust\_prospect\_type\_key#62,offer\_key#63,location\_key#64,agent\_key#65,nbr\_responses#66,nbr\_accept\_responses#67,nbr\_decline\_responses#68,nbr\_maybe\_later\_responses#69,nbr\_offers\_presented#70,nbr\_successful\_responses#71,nbr\_qualified\_offers#72,nbr\_qualified\_offers\_top\_3#73,nbr\_qualified\_offers\_nba#74,nbr\_qualified\_offers\_pos\_2#75,nbr\_qualified\_offers\_pos\_3#76,nbr\_nba\_presented#77,cif\_key#78,cif\_perm\_key#79,session\_id#80,offer\_qualification\_rank#81,offer\_display\_rank#82,nbr\_control\_group\_offers#83,nbr\_no\_offer\_offers#84,nbr\_expedite\_offers#85,card\_number#86,nbr\_qualified\_offers\_notofferd#87,sub\_disposition\_key#88,call\_key#89,branch\_id#90,employee\_id#91,bank#92,disposition\_key#93,device\_key#94,business\_segment\_key#95,operating\_system\_key#96,created\_date#97,hdfs\_load\_ts#98,date\_key#99] ParquetRelation[hdfs://bdp-isilonuat01.uat.pncint.net:8020/bdpd/bdh/01/str/pub/bdhhd01d\_bdh/cust\_intractn\_event\_fact]

== Physical Plan ==

TungstenAggregate(key=[], functions=[(count(1),mode=Final,isDistinct=false)], output=[count#120L])

TungstenExchange SinglePartition

TungstenAggregate(key=[], functions=[(count(1),mode=Partial,isDistinct=false)], output=[currentCount#123L])

TungstenProject

SortMergeOuterJoin [DATE\_KEY#39], [DATE\_KEY#99], LeftOuter, None

TungstenSort [DATE\_KEY#39 ASC], false, 0

TungstenExchange hashpartitioning(DATE\_KEY#39)

ConvertToUnsafe

Scan ParquetRelation[hdfs://bdp-isilonuat01.uat.pncint.net:8020/bdpd/bdh/01/str/pub/bdhhd01d\_bdh/cust\_intractn\_event\_fact/date\_key=20160822][DATE\_KEY#39]

TungstenSort [DATE\_KEY#99 ASC], false, 0

TungstenExchange hashpartitioning(DATE\_KEY#99)

ConvertToUnsafe

Scan ParquetRelation[hdfs://bdp-isilonuat01.uat.pncint.net:8020/bdpd/bdh/01/str/pub/bdhhd01d\_bdh/cust\_intractn\_event\_fact][DATE\_KEY#99]