**== Physical Plan ==**

**CollectLimit (8)**

**+- \* HashAggregate (7)**

**+- Exchange (6)**

**+- \* HashAggregate (5)**

**+- Exchange (4)**

**+- \* Project (3)**

**+- \* Filter (2)**

**+- Scan csv (1)**

**(1) Scan csv**

**Output [2]: [age#21, country#22]**

**Batched: false**

**Location: InMemoryFileIndex [dbfs:/FileStore/tables/sample\_data.csv]**

**PushedFilters: [IsNotNull(age), LessThan(age,40)]**

**ReadSchema: struct<age:int,country:string>**

**Jobid 2 has (2,3,4 stages)**

**Stages 2 has (scan-> codegen->excahnge) 3 taskes**

**Skip 0**

**File size 256.0 B**

**Read 0**

**Write 326.0 B**

**Partition 1**

**(2) Filter [codegen id : 1]**

**Input [2]: [age#21, country#22]**

**Condition : (isnotnull(age#21) AND (age#21 < 40))**

**Jobid 2 has (2,3,4 stages)**

**Stages 2 has (scan-> codegen->excahnge)**

**Skip 0**

**File size 256.0 B**

**Read 0**

**Write 326.0 B**

**Partition 1**

**(3) Project [codegen id : 1]**

**Output [1]: [country#22]**

**Input [2]: [age#21, country#22]**

**Jobid 2 has (2,3,4 stages)**

**Stages 2 has (scan-> codegen->excahnge)**

**Skip 0**

**File size 256.0 B**

**Read 0**

**Write 326.0 B**

**Partition 1**

**(4) Exchange**

**Input [1]: [country#22]**

**Arguments: RoundRobinPartitioning(4), REPARTITION\_BY\_NUM, [plan\_id=47]**

**Jobid 2 has (2,3,4 stages)**

**Stages 2(scan-> codegen->excahnge) ,3(excahnge->codegen->excahnge) //repartion(4)**

**Skip**

**File size 326.0 B**

**Read -> 3-stage-> 326.0 B**

**Write -> 2-stage-> 326.0 B , 3-stage->506 B**

**Partition 2-stage->1 and 3-stage-> 4**

**//we have 4 files based on country as key**

**p1->canada data**

**p2->Us**

**p4**

**(5) HashAggregate [codegen id : 2]**

**Input [1]: [country#22]**

**Keys [1]: [country#22]**

**Functions [1]: [partial\_count(1) AS count#52L]**

**Aggregate Attributes [1]: [count#51L]**

**Results [2]: [country#22, count#52L]**

**Jobid 2**

**Stages 3(excahnge->codegen->excahnge)**

**Skip**

**File size 326.0 B**

**Read 3-stage-> 326.0 B**

**Write 3-stage->506 B**

**Partition 3-stage-> 4**

**//we have 4 files based on country as key**

**p1->canada data**

**p2->Us**

**p4**

**(6) Exchange**

**Input [2]: [country#22, count#52L]**

**Arguments: hashpartitioning(country#22, 200), ENSURE\_REQUIREMENTS, [plan\_id=51]**

**Jobid 2,3,4,5,6**

**Stages** [**3.0**](https://community.cloud.databricks.com/sparkui/1022-090019-1ntfok75/driver-521416917290199348/stages/stage/?id=3&attempt=0)[**4.0**](https://community.cloud.databricks.com/sparkui/1022-090019-1ntfok75/driver-521416917290199348/stages/stage/?id=4&attempt=0)[**7.0**](https://community.cloud.databricks.com/sparkui/1022-090019-1ntfok75/driver-521416917290199348/stages/stage/?id=7&attempt=0)[**10.0**](https://community.cloud.databricks.com/sparkui/1022-090019-1ntfok75/driver-521416917290199348/stages/stage/?id=10&attempt=0)[**13.0**](https://community.cloud.databricks.com/sparkui/1022-090019-1ntfok75/driver-521416917290199348/stages/stage/?id=13&attempt=0)[**16.0**](https://community.cloud.databricks.com/sparkui/1022-090019-1ntfok75/driver-521416917290199348/stages/stage/?id=16&attempt=0)

**Skip 3,4,5,6**

**File size 506 B**

**Read**

**3-stage ->326B ,-stage 4→0**

**,7-stage ->0 ,10-stage->142B ,13-stage->142B ,14-stage->222B**

**Write**

**3-stage ->506B ,-stage 4->0**

**,7-stage ->0 ,10-stage->0 ,13-stage->0 ,14-stage->0**

**Partiton**

**3-stage ->4 ,stage -4->1**

**,7-stage ->4 ,10-stage->20 ,13-stage->100 ,14-stage->75**

**200 final partiton**

**(7) HashAggregate [codegen id : 3]**

**Input [2]: [country#22, count#52L]**

**Keys [1]: [country#22]**

**Functions [1]: [finalmerge\_count(merge count#52L) AS count(1)#41L]**

**Aggregate Attributes [1]: [count(1)#41L]**

**Results [2]: [country#22, cast(count(1)#41L as string) AS count#48]**

**Jobid**

**Stages** [**4.0**](https://community.cloud.databricks.com/sparkui/1022-090019-1ntfok75/driver-521416917290199348/stages/stage/?id=4&attempt=0)[**7.0**](https://community.cloud.databricks.com/sparkui/1022-090019-1ntfok75/driver-521416917290199348/stages/stage/?id=7&attempt=0)[**10.0**](https://community.cloud.databricks.com/sparkui/1022-090019-1ntfok75/driver-521416917290199348/stages/stage/?id=10&attempt=0)[**13.0**](https://community.cloud.databricks.com/sparkui/1022-090019-1ntfok75/driver-521416917290199348/stages/stage/?id=13&attempt=0)[**16.0**](https://community.cloud.databricks.com/sparkui/1022-090019-1ntfok75/driver-521416917290199348/stages/stage/?id=16&attempt=0)

**Skip**

**File size 506 B**

**Read**

**stage 4→0**

**,7-stage ->0 ,10-stage->142B ,13-stage->142B ,14-stage->222B**

**Write**

**-stage 4->0**

**,7-stage ->0 ,10-stage->0 ,13-stage->0 ,14-stage->0**

**Partition**

**stage -4->1**

**,7-stage ->4 ,10-stage->20 ,13-stage->100 ,14-stage->75**

**(8) CollectLimit**

**Input [2]: [country#22, count#48]**

**Arguments: 21**

**Jobid**

**Stages**

**Skip**

**File size**

**Read**

**write**