Apache Spark

Execution Plan

Spark Execution Plan

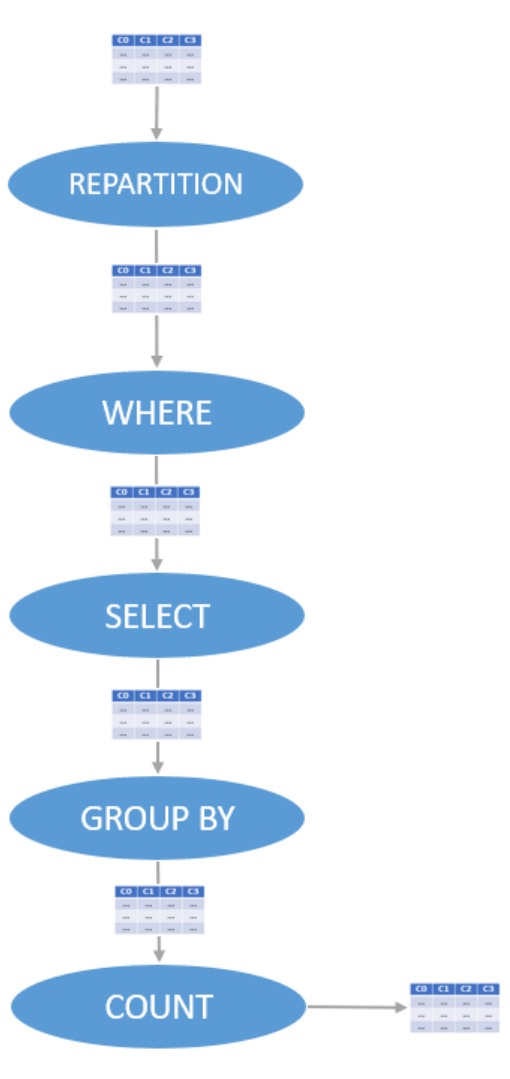
```scala
val surveyRawDF = spark.read
  .option("header", "true")
  .option("inferSchema", "true")
  .csv(args(0))
```

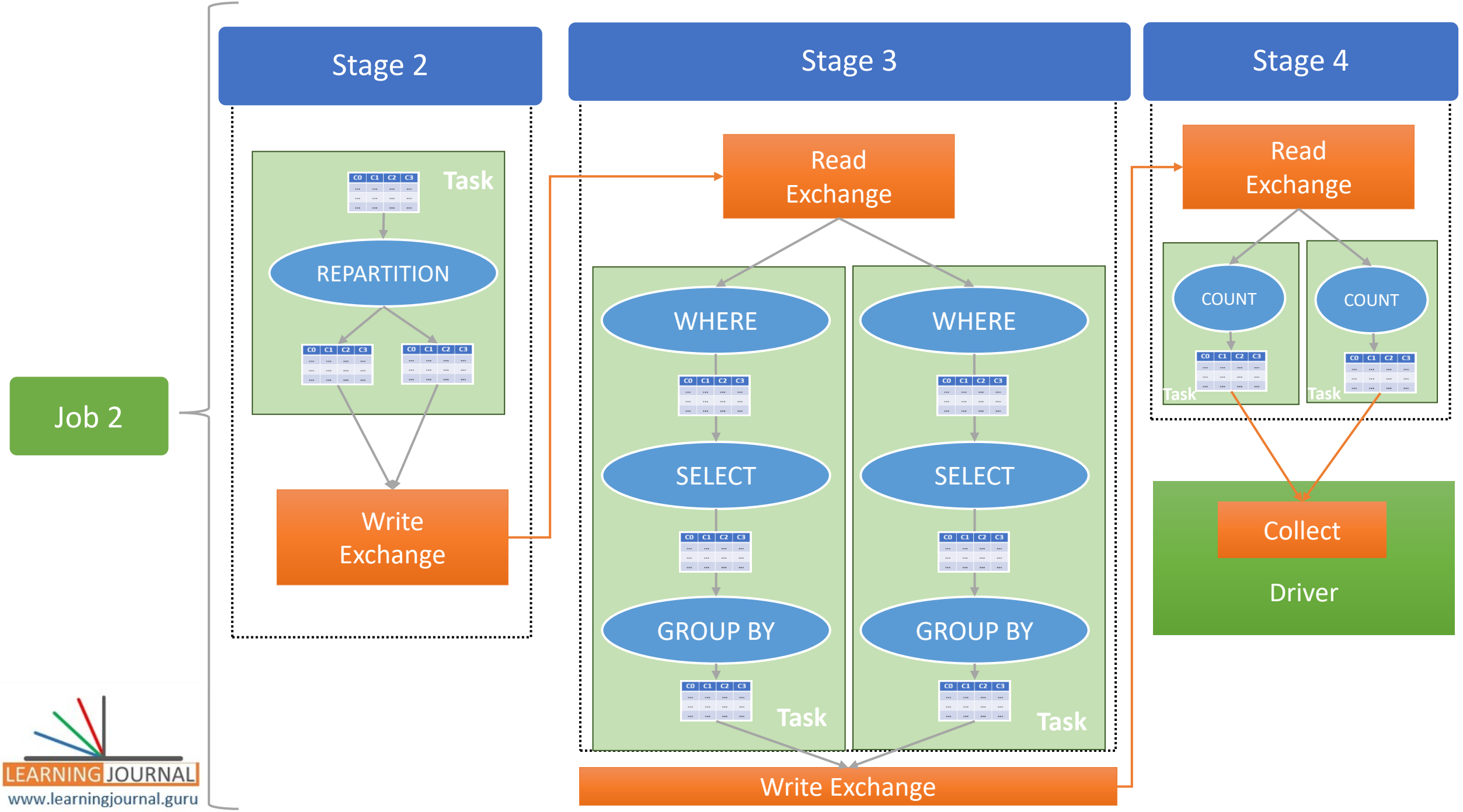
Job 0

Job 1

```scala
val partitionedSurveyDF = surveyRawDF.repartition( numPartitions = 2)
val countDF = partitionedSurveyDF.where( conditionExpr = "Age < 40")
  .select( col = "Age",  cols = "Gender", "Country", "state")
  .groupBy( col1 = "Country")
  .count()
logger.info(countDF.collect().mkString("->"))
```

Job 2




Job 2
Stage 2
Task
REPARTITION
Write Exchange
Stage 3
Read Exchange
WHERE
SELECT
GROUP BY
Task
WHERE
SELECT
GROUP BY
Task
Write Exchange
Stage 4
Read Exchange
COUNT
COUNT
Task
Task
Collect
Driver
LEARNING JOURNAL
www.learningjournal.guru


LEARNING JOURNAL
www.learningjournal.guru