Spark Skew Join and Handling It.

 the idea here is to divide larger partitions into smaller ones using “salt” (our own created new column) but it also comes with side effect of getting smaller partitions divided into even more smaller ones.

Property :

To use in spark-sql:

Spark.udf.register(“name to be”, function\_name,StringType())

To use in spark api :

Udf\_name\_to = udf(lambda z: function\_name(z))

SparkContext. setCheckpointDir(directory: String).

accum=sc.accumulator(0)

rdd=spark.sparkContext.parallelize([1,2,3,4,5])

rdd.foreach(lambda x:accum.add(x))

print(accum.value) #Accessed by driver

val mapDF=df.map(fun=> {

fun.getString(0).split(" ")

})