

KYLIN-5441 The assignment logic of the spark.sql.shuffle.partitions parameter is incorrect

Description

A query will cut out multiple OlapContext, each context will call ResetShufflePartition#setShufflePartitions method once, in this method will calculate a partitionsNum based on the number of scanned bytes (SourceScanBytes) and assigns this value to spark.sql.shuffle.partitions.

The problem is that this assignment will overwrite the previous assignment, which is not reasonable because there is a possibility that a smaller value will overwrite a larger value, reducing the number of partitions in the shuffle and leading to a drop in query performance. The expectation is to keep the maximum partitionNum.

Fix design

Add the shufflePartitionsReset field to the QueryContext to record the reset shuffle partition
Based on the shufflePartitionsReset field to determine whether to carry out the shuffle partitions reset

Description

一个查询会切出多个OlapContext，每个context都会调用一次 ResetShufflePartition#setShufflePartitions方法，在这个方法里会根据扫描字节数 (SourceScanBytes)计算出一个 partitionsNum，并将这个数值赋值给 spark.sql.shuffle.partitions。

问题在于这样赋值会把前次赋值给覆盖掉，这是不合理的，因为有可能出现较小的数值覆盖较大的数值，降低shuffle时的partition数，导致查询性能下降。期望保留最大的partitionNum。

Fix design

在QueryContext 中添加 `shufflePartitionsReset` 字段，用来记录reset 的shuffle partition
根据 `shufflePartitionsReset` 这个字段来判断是否进行 shuffle partitions 的reset