### 一、背景：

-- 452264

    -- 处理航空件时，聚合时，并行度太低。Task1=37 ，Task2 = 17  ，是否可参考 第一步聚合，关闭推测执行

        -- 1、查看DAG, 此步骤走AQE，导致并行度 为 37,17  ，考虑关闭并行度。

        -- 2、数据量

            输入表： 1374542822

            select count(1) from tmp\_ordi\_predict.tmp\_shenhy\_air\_special\_econ\_tmp4

            输出表： 7058551

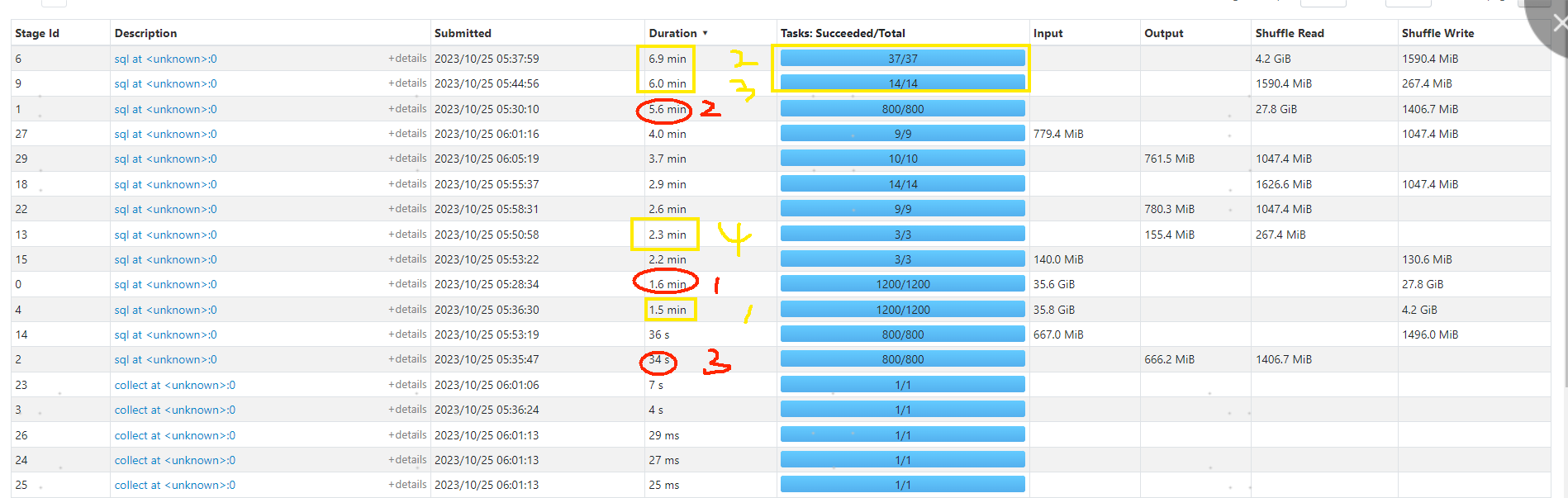
            select count(1) from tmp\_ordi\_predict.tmp\_shenhy\_air\_special\_econ\_tmp6;

        -- 3、测试applicationID ： application\_1695122323486\_17191600

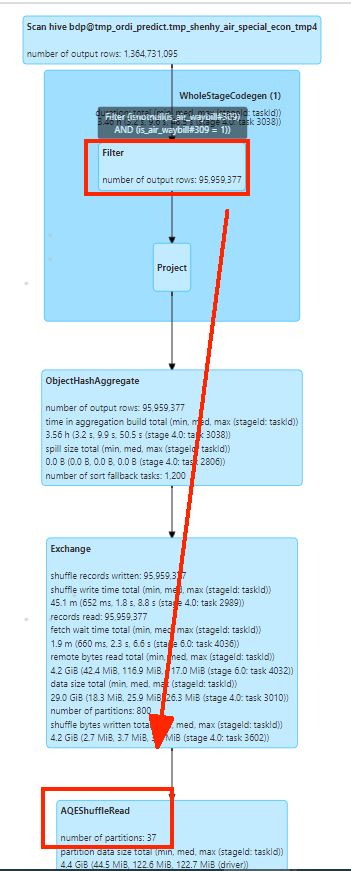
### 二、问题定位：

#### Stage并行度查看（查看历史耗时最长的application\_1695122323486\_17191600）

**黄色1,2,3**为第二步过滤航空件的聚合，相比较**红色1,2,3（已关闭AQE）**并行度低很多



#### 查看DAG（application\_1695122323486\_17191600）



### 三、优化措施：

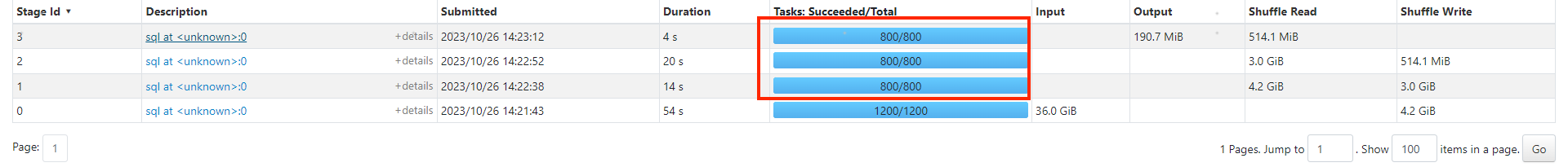
#### 1、参数调整

set spark.sql.shuffle.partitions=800;  -- 增加并行度

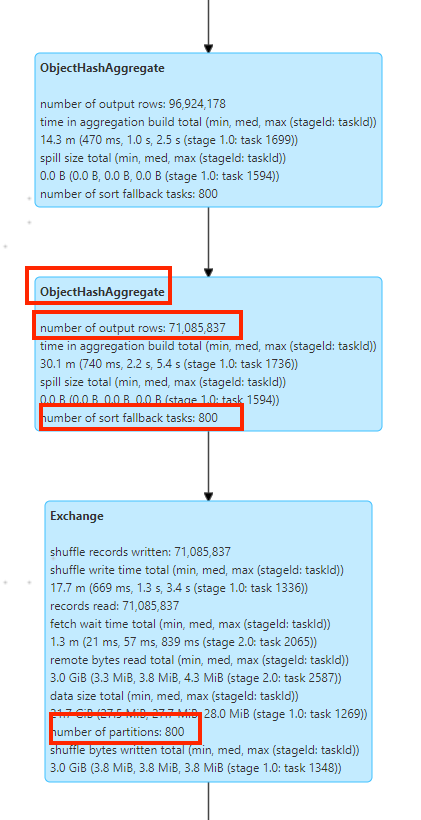
set spark.sql.adaptive.enabled=false;  -- 临时关闭适配执行

#### 2、查看stage并行度

并行度增加至800，速度快了



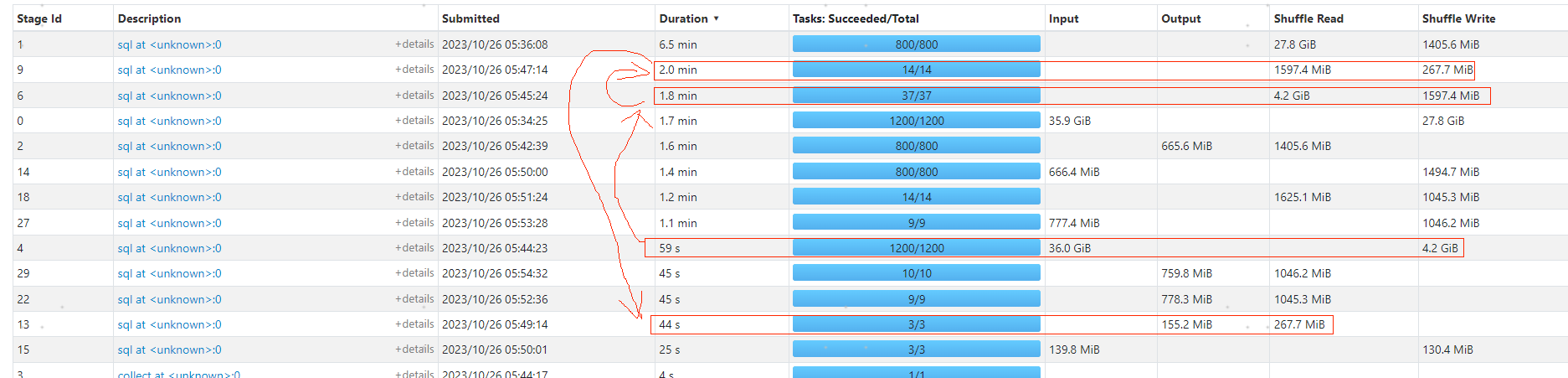
#### 查看DAG



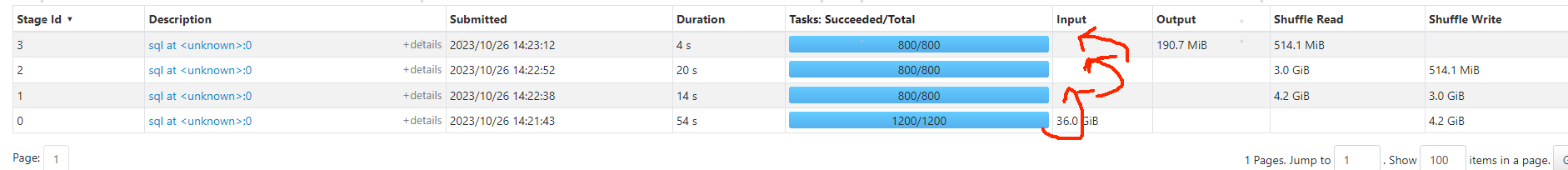
### 优化前后耗时对比

1. stage读取耗时对比（数据量一样对比）

前：application\_1695122323486\_17022940



后：application\_1695122323486\_17191600



2-、耗时对比

