Requirement is to transform the data using spark, write the transformed data into delta table and capture the count to insert into audit table.  
  
How would you optimize above ask ??  
  
Basic optimization technique - Since we are calling two actions ( write & count) on a same data frame, data frame can be cached or persisted.  
  
Cache will have performance issues if the data size is huge and it has to go through each and every record in memory to get the count.  
  
if the target table created with delta format, make use of below code to get the count from delta log operation metrics.  
  
Select OPERATIONMETRICS.numoutputRows as count from (Describe history tableName order by version desc limit 1)  
  
I have tried with 3 approaches :  
  
Writing the data + getting the count without cache - 26.99 mins  
  
Writing the data + getting the count with cache - 19.76 mins  
  
Writing the data + getting the count from operation metrics - 15.83 mins

