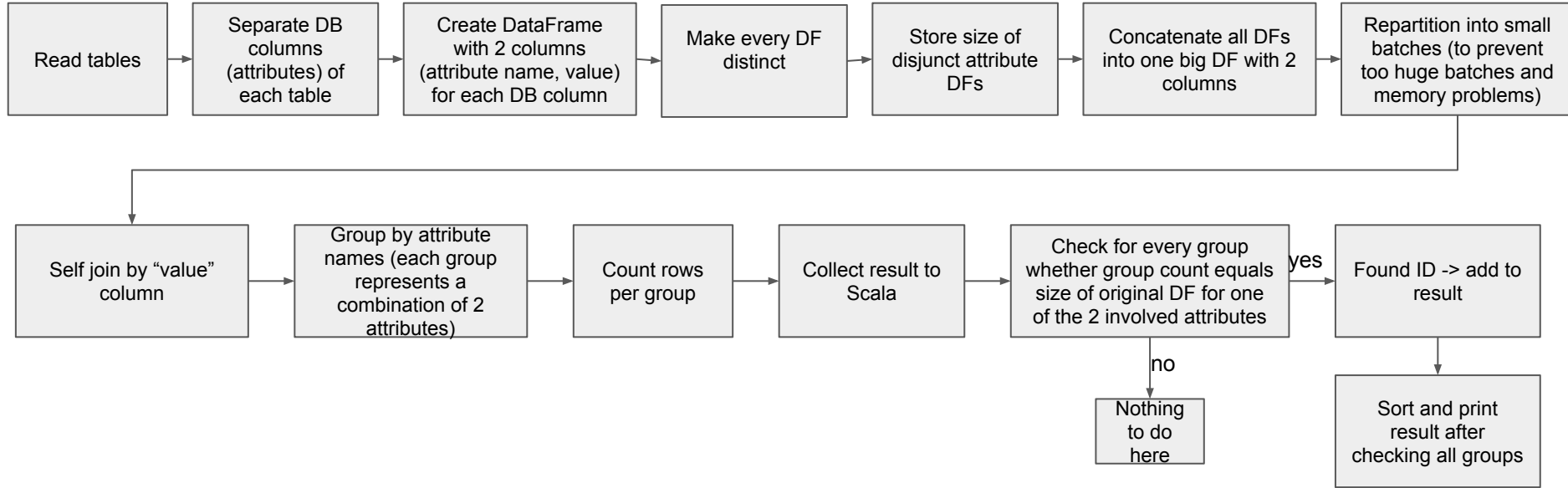


Distributed Data Exercise 2

Task 3, Team MeMyselfAndSomeoneElse

General Workflow



Remarks:

- Set spark configuration *spark.driver.maxResultSize* to "8g" to prevent memory problems with limited performance Laptop which we used for development