Patrick Porter

Project Retrospective P3

1. MapReduce
    Pros - Highly scalable
    Cons - Restrictive
  Spark
    Pros - Highly scalable and less restrictive than MapReduce
    Cons - Requires a lot of RAM
2. Anything where you're comparing data across different files in HDFS is more difficult that I thought.
3. User defined functions made things easier. I originally thought I could use the map function and lambdas with DataFrames but quickly learned that wasn't the case.
4. Roughly 15 hours. Getting familiar with
5. I learned the basics of working with Spark.
6. I started too late! Didn't get to come up with as detailed answers to the questions as I would have liked.