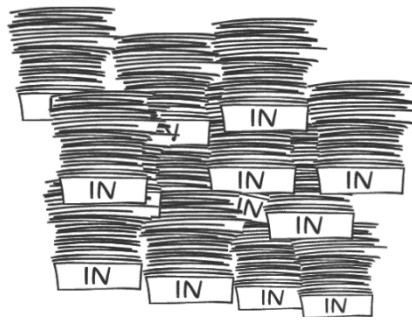


In the beginning there was...

All of this big data...
What am I supposed
to do with it?



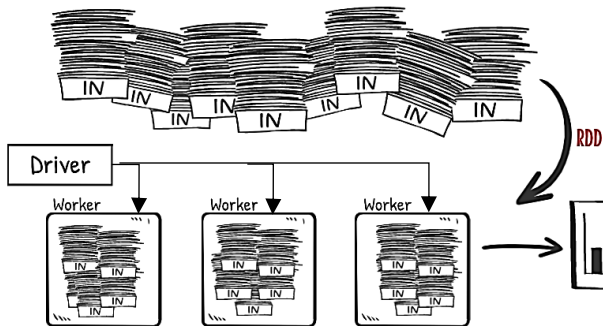
...So Scholars gathered to find a solution...

AGENDA

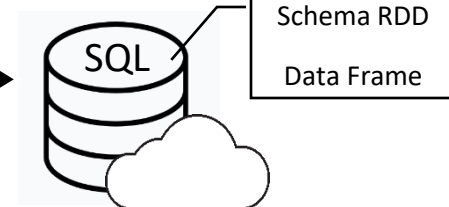
1. MLlib
2. SQL
3. Streaming



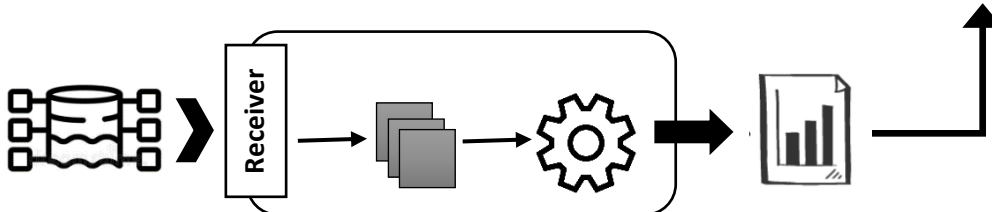
1. MLlib



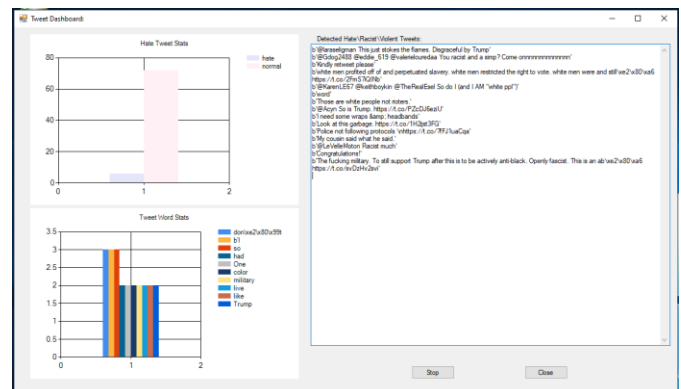
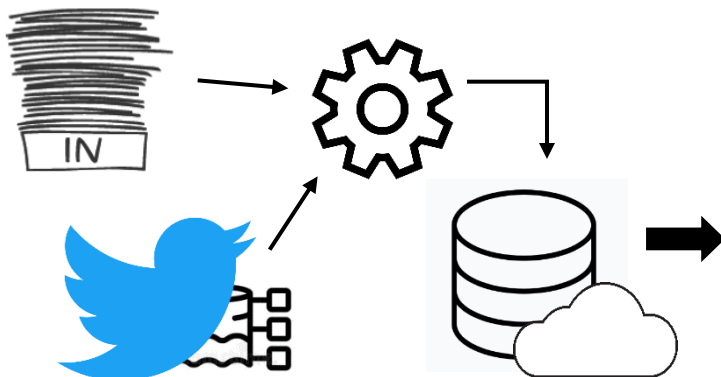
2. SQL



3. Streaming



...And at last a test was in order... **Twitter Analysis**



RDD RESILIENT DISTRIBUTED DATASETS

SPARK ENGINE

BATCHES