1. Introduction
2. Input Data
3. Input Data
4. Defensive Mapper Code
5. Defensive Mapper Code
6. Between (mapper and reducer)
   1. Uses shuffle and sort
7. Reducing
   1. Hadoop streaming = allows hadoop to use any language
   2. Keeping track

VIII. Reducing

IX. Reducer Code

X. Reducer Code

XI. Putting all together:

* Can run without hadoop for testing. Simply run the code.
* Mapper: ***python mapper.py*** or create a sample data file and feed to mapper i.e ***cat testfile | python mapper.py***
* Reducer: ***cat testfile | ./mapper.py | sort | ./reducer.py*** # simulates hadoop by sorting
* Hadoop: ***hs mapper.py reducer.py myinput output2***
  + - ***hadoop fs -ls output2***
    - ***hadoop fs -cat output2/part-00000***
* Results and ongoing processes can be seen in ***localhost:50030/jobtracker.jsp***

XII. Conclusion