

Lab 4: Designing and Tuning MapReduce Algorithms

CS 660 Data Science at Scale
Jeremy Johnson

Reference

- ❖ The lab refers to the following book
- ❖ Available through Week 4 Resources in BBLearn and online
- ❖ Reference
 - Jimmy Lin and Chris Dyer, Data-Intensive Text Processing with MapReduce, Morgan & Claypool Publishers, 2010.

Lab 4 (MapReduce Patterns)

1. Review chapters 2 and 3 in MapReduce book.
2. Read and summarize chapter 4 on Inverted Indexing for Text Retrieval
3. Implement, using MRJob, the algorithms presented for inverted indexing.
4. Run and measure performance on Google Cloud with varying number of nodes