Reading list for Hadoop
(in the context of Digital Preservation/Digital Libraries)

- Tom White
  Hadoop. The Definitive Guide

- Tom White
  The small files problem

- Jimmy Lin
  MapReduce is Good Enough? If All You Have is a Hammer, Throw Away Everything That's Not a Nail!

- Jimmy Lin and Chris Dyer
  Data-Intensive Text Processing with MapReduce

- Jimmy Lin, Milad Gholami, and Jinfeng Rao
  Infrastructure for Supporting Exploration and Discovery in Web Archives

- Mohamed Mohideen Abdul Rasheed
  Fedora Commons With Apache Hadoop: A Research Study

- Frank Asseg, Matthias Razum and Matthias Hahn
  Apache Hadoop as a Storage Backend for Fedora Commons

- Martin Rennhackkamp
  Big Data, BI and Digital Preservation in Cloud-based Storage

- Rainer Schmidt
  An Architectural Overview of the SCAPE Preservation Platform

- Rainer Schmidt and Matthias Rella
  An Approach for Processing Large and Non-Uniform Media Objects on MapReduce-based Clusters

- Peter Wittek and Sándor Darányi
  Leveraging on High-Performance Computing and Cloud Technologies in Digital Libraries: A Case Study

- Orit Edelstein, Michael Factor, Ross King, Thomas Risse, Eliot Salant and Philip Taylor
  Evolving Domains, Problems and Solutions for Long Term Digital Preservation