CS744: Big Data Systems

Jack Truskowski

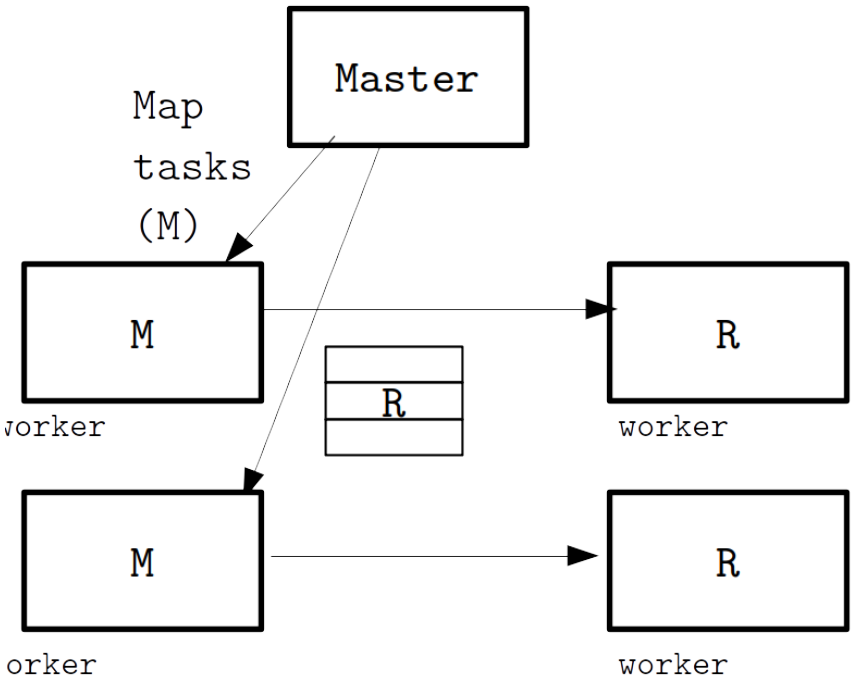
February 4, 2019

Contents

1 2/4/19 MapReduce **1**
1.1 Operators 2
1.2 Failures and Slowdowns 2
1.2.1 Possible failures 3

1 2/4/19 MapReduce

- Programming model
- Execution
- Runtime issues
- M-R library handles execution and run-time issues
  - Transparent to programmers



1.1 Operators

1. Map
   - Input = (key,value) –> (key, <v>)
2. Reduce
   - Operates share a key
   - (key,value) is sorted and values passed to reducer

1.2 Failures and Slowdowns

- Handled by the master

1.2.1 Possible failures

1. Map / Reduce
    - Worker fails, some maps and some reduces completed
    - Reduce data is already written to HDFS, doesn't need to be re-computed
    - Maps must be re-executed, since it hasn't been written to HDFS