

CSC 369: Distributed Computing

Alex Dekhtyar

May 8

Day 15: Java Hadoop API and Examples







Housekeeping

Lab 4 (mini-project): due Sunday night

Lab 5: you should be done with it

Lab 6: Out. Due Monday, May 18

Lab 7: Mini-project 2, assigned at the end of next week.

Labs 8,9 - short Spark labs

Housekeeping: Midterm 1

10 problems, 70 points reachable

Mean	24.125
Median	22.5
St.Dev	10.828
Max	52
Min	6.5

6.5	22.5
6.5	23
12	25
13	27.5
14	28
14	29
14.5	29.5
16.5	33
19	34
19.5	35.5
19.5	36
20.5	39.5
22	41
22.5	52

Hadoop Java API

Hadoop Java API

`org.apache.hadoop`

`org.apache.hadoop.mapreduce`

Core MapReduce classes

`org.apache.hadoop.mapreduce.lib.input`

`org.apache.hadoop.mapreduce.lib.output`



Inuput/Output
parsing

`org.apache.hadoop.io`

atomic type wrappers

`org.apache.hadoop.conf`

Job configuration

`org.apache.hadoop.fs`

File system classes

`org.apache.hadoop.mapreduce`

`org.apache.hadoop.mapreduce.Job` MapReduce Job

`org.apache.hadoop.mapreduce.Mapper` Extensible Mapper

`org.apache.hadoop.mapreduce.Reducer` Extensible Reducer

`org.apache.hadoop.mapreduce.Partitioner` Parent class for Partitioning tasks

`org.apache.hadoop.mapreduce.InputFormat` Parent classes for Input/Output Formats
`org.apache.hadoop.mapreduce.OutputFormat`

`org.apache.hadoop.mapreduce.InputSplit` Parent class for Input Split

`org.apache.hadoop.mapreduce.Job`

`org.apache.hadoop.io`

Demos

1. Splitting inputs, multiplying reducers (NLGroup.java)
2. Multiple MapReduce Jobs (filter.java)
3. Multiple Input Files (multiInMR.java)