

# MapReduce Homework 2 - Report

nju\_st 42

洪亮(181240019), 孙舒禹(181840204), 曾许墨秋(181240004)

18匡院数理 NJU

## Map Reduce 设计思路

### jobInvertedIndex

- function: 计算实验要求的倒排索引
- output format: Word \t Freq, (Filename: Occurrence;)\*n
- Mapper<LongWritable, Text, Text, Text>: get filename, for each word, output (key: word, value: filename)
- Reducer<Text, Text, Text, Text>: Using two vector to record filename and occurrences for each file, output result

### jobConvertFreqToKey

- function: 提取Freq作为key, word 作为value, 存在outputpath/temp/ConvertFreqToKey中, 作为中间文件
- input format: setInputFormatClass(KeyValueTextInputFormat.class), token with \t
- output format: Freq \t Word
- Mapper<Text, Text, Text, Text>: token word and freq, output (key: freq, value: word)
- Reducer<Text, Text, Text, Text>: do nothing

### jobSorted

- function: Sort InvertedIndex result by freq, using TotalOrderPartitioner
- input format: use output of jobConvertFreqToKey, setInputFormatClass(KeyValueTextInputFormat.class), token with \t
- output format: Word \t Freq
- Comparator: change Text -> Sting -> Float, then cal Float.compare to compare
- Mapper<Text, Text, Text, Text>: do nothin
- Reducer<Text, Text, Text, Text>: exchange key and value

## Code with commend

```
import com.ibm.jvm.j9.dump.extract.Main;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.*;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.FileSplit;
import org.apache.hadoop.mapreduce.lib.input.KeyValueTextInputFormat;
import org.apache.hadoop.mapreduce.lib.jobcontrol.JobControl;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
```

```

import org.apache.hadoop.mapreduce.lib.partition.InputSampler;
import org.apache.hadoop.mapreduce.lib.partition.TotalOrderPartitioner;
import org.apache.hadoop.util.GenericOptionsParser;

import java.io.IOException;
import java.net.URI;
import java.util.Arrays;
import java.util.Iterator;
import java.util.StringTokenizer;
import java.util.Vector;

public class MainClass {
    public MainClass() {}

    public static void main(String[] args) throws Exception {
        Configuration conf = new Configuration();
        String[] otherArgs = (new GenericOptionsParser(conf,
args)).getRemainingArgs();
        if (otherArgs.length != 3) {
            System.err.println(otherArgs[0]);
            System.err.println("Usage: MainClass <in> <out>");
            System.exit(2);
        }

        // Setting jobInvertedIndex
        Job jobInvertedIndex = Job.getInstance(conf, "Invert Index");
        jobInvertedIndex.setJarByClass(MainClass.class);
        jobInvertedIndex.setMapperClass(MainClass.InvertedIndexMapper.class);
        jobInvertedIndex.setReducerClass(MainClass.InvertedIndexReducer.class);
        jobInvertedIndex.setOutputKeyClass(Text.class);
        jobInvertedIndex.setOutputValueClass(Text.class);
        // Setting input & output dictionary for jobInvertedIndex
        FileInputFormat.addInputPath(jobInvertedIndex, new Path(otherArgs[1]));
        FileOutputFormat.setOutputPath(jobInvertedIndex, new
Path(otherArgs[2]+"/InvertedIndex"));
        // Execute jobInvertedIndex
        jobInvertedIndex.waitForCompletion(true);

        // Setting jobConvertFreqToKey
        Configuration convertFreqToKeyConf = new Configuration();

        convertFreqToKeyConf.set("mapreduce.input.keyvaluelinerecordreader.key.value.se
parator", "\\t");
        Job jobConvertFreqToKey = Job.getInstance(convertFreqToKeyConf, "Convert
Freq to Key");
        jobConvertFreqToKey.setInputFormatClass(KeyValueTextInputFormat.class);
        jobConvertFreqToKey.setJarByClass(MainClass.class);

        jobConvertFreqToKey.setMapperClass(MainClass.ConvertFreqToKeyMapper.class);

        jobConvertFreqToKey.setReducerClass(MainClass.ConvertFreqToKeyReducer.class);
        // Setting type for jobConvertFreqToKey
        jobConvertFreqToKey.setMapOutputKeyClass(Text.class);
        jobConvertFreqToKey.setMapOutputValueClass(Text.class);
        jobConvertFreqToKey.setOutputKeyClass(Text.class);
        jobConvertFreqToKey.setOutputValueClass(Text.class);
        // Setting path for jobConvertFreqToKey
    }
}

```

```

        FileInputFormat.addInputPath(jobConvertFreqToKey, new
Path(otherArgs[2]+"/InvertedIndex"));
        FileOutputFormat.setOutputPath(jobConvertFreqToKey, new
Path(otherArgs[2]+"/temp/ConvertFreqToKey"));
        // Execute jobConvertFreqToKey
        jobConvertFreqToKey.waitForCompletion(true);

        // Setting jobSorted
        Configuration sortedConf = new Configuration();

        sortedConf.set("mapreduce.input.keyvaluelinerecordreader.key.value.separator",
"\t");
        Job jobSorted = Job.getInstance(sortedConf, "Sorted");
        jobSorted.setInputFormatClass(KeyValueTextInputFormat.class);
        jobSorted.setNumReduceTasks(4);
        // setting types for jobSorted
        jobSorted.setMapOutputKeyClass(Text.class);
        jobSorted.setMapOutputValueClass(Text.class);
        jobSorted.setOutputKeyClass(Text.class);
        jobSorted.setOutputValueClass(Text.class);
        // Setting path for jobSorted
        FileInputFormat.addInputPath(jobSorted, new
Path(otherArgs[2]+"/temp/ConvertFreqToKey"));
        FileOutputFormat.setOutputPath(jobSorted, new
Path(otherArgs[2]+"/Sorted"));
        // Setting Class
        jobSorted.setJarByClass(MainClass.class);
        jobSorted.setMapperClass(MainClass.SortedMapper.class);
        jobSorted.setReducerClass(MainClass.SortedReducer.class);
        // Setting partitioner & comparator
        jobSorted.setSortComparatorClass(SortedKeyComparator.class);
        TotalOrderPartitioner.setPartitionFile(jobSorted.getConfiguration(), new
Path(otherArgs[2]+"/temp/partition"));
        InputSampler.Sampler<Text, Text> sortedSampler = new
InputSampler.RandomSampler<>(0.01, 1000, 100);
        InputSampler.writePartitionFile(jobSorted, sortedSampler);
        jobSorted.setPartitionerClass(TotalOrderPartitioner.class);
        String partitionFile =
TotalOrderPartitioner.getPartitionFile(jobSorted.getConfiguration());
        URI uri = new URI(partitionFile+"#"+TotalOrderPartitioner.DEFAULT_PATH);
        jobSorted.addCacheFile(uri);
        // Execute jobSorted
        System.exit(jobSorted.waitForCompletion(true) ? 0 : 1);
    }

    // InvertedIndex mapper & reducer
    public static class InvertedIndexMapper extends Mapper<LongWritable, Text,
Text, Text> {
        private static final LongWritable one = new LongWritable(1);
        private Text word = new Text();

        public InvertedIndexMapper() {}

        public void map(LongWritable key, Text value, Mapper<LongWritable, Text,
Text, Text>.Context context) throws IOException, InterruptedException {
            // default: key - lineoffset, value - line
            StringTokenizer itr = new StringTokenizer(value.toString());
            FileSplit filesplit = (FileSplit)context.getInputSplit();

```

```

        Text fileName = new Text(fileSplit.getPath().getName());
        while (itr.hasMoreTokens()) {
            this.word.set(itr.nextToken());
            context.write(this.word, fileName);
        }
    }
}

public static class InvertedIndexReducer extends Reducer<Text, Text, Text,
Text> {
    private Vector<String> fileList = new Vector<String>();
    private Vector<Integer> occurrences = new Vector<Integer>();
    private String resultString = "";
    private Text resultText = new Text();

    public InvertedIndexReducer() {}

    public void reduce(Text key, Iterable<Text> values, Reducer<Text, Text,
Text, Text>.Context context) throws IOException, InterruptedException {
        // key: word, values: filenames
        String currentFileName;
        for(Iterator itr = values.iterator(); itr.hasNext();) {
            currentFileName = itr.next().toString();
            if (!fileList.contains(currentFileName)) { // if new file, add
to list
                fileList.add(currentFileName);
                occurrences.add(0);
            }
            // update occurrence num
            int currentIndex = fileList.indexOf(currentFileName);
            occurrences.set(currentIndex, occurrences.get(currentIndex)+1);
        }

        // output results
        int sumOccurrence = 0;
        int fileNumber = fileList.size();
        String appendValue = "";
        for (int i = 0; i < fileNumber; i++) {
            sumOccurrence += occurrences.get(i);
            appendValue += (fileList.get(i) + ":" + occurrences.get(i));
            if (i!=(fileNumber-1)) {
                appendValue += "; ";
            }
        }
        float averageOccurrence = (float)sumOccurrence / (float)fileNumber;
        this.resultString = String.format("%.2f", averageOccurrence) + ", "
+ appendValue;
        this.resultText.set(resultString);
        context.write(key, this.resultText);
    }
}

// ConvertFreqToKey mapper & reducer
public static class ConvertFreqToKeyMapper extends Mapper<Text, Text, Text,
Text> {

    public ConvertFreqToKeyMapper() {}

```

```

        public void map(Text key, Text value, Mapper<Text, Text, Text,
Text>.Context context) throws IOException, InterruptedException {
            // key - key, value - value (token by '\t')
            StringTokenizer inputValueTokenizer = new
StringTokenizer(value.toString());
            String outputKeyString = inputValueTokenizer.nextToken(",");    //
token freq
word
            context.write(new Text(outputKeyString), key); // output: freq\t
        }
    }

    public static class ConvertFreqToKeyReducer extends Reducer<Text, Text,
Text, Text> {

        public ConvertFreqToKeyReducer() {}

        public void reduce(Text key, Iterable<Text> values, Reducer<Text, Text,
Text, Text>.Context context) throws IOException, InterruptedException {
            // key: word, values: filenames
            for (Text value: values) { // do nothing
                context.write(key, value);
            }
        }
    }

    // Sorted mapper & reducer & comparator
    public static class SortedMapper extends Mapper<Text, Text, Text, Text> {

        public SortedMapper() {}

        public void map(Text key, Text value, Mapper<Text, Text, Text,
Text>.Context context) throws IOException, InterruptedException {
            // key - key, value - value
            // do nothing
            context.write(key, value);
        }
    }

    public static class SortedReducer extends Reducer<Text, Text, Text, Text> {

        public SortedReducer() {}

        public void reduce(Text key, Iterable<Text> values, Reducer<Text, Text,
Text, Text>.Context context) throws IOException, InterruptedException {
            // key: word, values: filenames
            for (Text value: values) { // switch key and values
                context.write(value, key);
            }
        }
    }

    public static class SortedKeyComparator extends WritableComparator {
        protected SortedKeyComparator() {
            super(Text.class, true);
        }
    }

```

```

@Override
public int compare(WritableComparable writableComparable1,
WritableComparable writableComparable2) {
    float float1 = Float.parseFloat(writableComparable1.toString());
    float float2 = Float.parseFloat(writableComparable2.toString());

    return -Float.compare(float1, float2);
}
}
}

```

## Output Result on Hdfs

HDFS分布式存储 模拟超级用户

[上传](#) [新建](#) [操作](#) [移动到回收站](#)

[根目录](#) /user/nju\_st42/ [Trash](#)

<input type="checkbox"/>	文件名	大小	用户	用户组	权限	上次修改
<input type="checkbox"/>	output	0B	nju_st42	supergroup	drwxr-xr-x	Wed May 05 2021 13:06:29 GMT+0800 (中国标准时间)
<input type="checkbox"/>	output_20210507_10_56	0B	nju_st42	supergroup	drwxr-xr-x	Fri May 07 2021 10:55:27 GMT+0800 (中国标准时间)
<input type="checkbox"/>	output_20210507_13_32	0B	nju_st42	supergroup	drwxr-xr-x	Fri May 07 2021 13:31:04 GMT+0800 (中国标准时间)
<input type="checkbox"/>	output_20210507_23_09	0B	nju_st42	supergroup	drwxr-xr-x	Fri May 07 2021 23:07:40 GMT+0800 (中国标准时间)

Show 10 of 4 items Page 1 of 1

图1 File Path on Hdfs

如图一三个output文件夹分别是三次测试结果，后缀是运行时间。

- output内只包含倒排索引
- output\_20210507\_10\_56内除了倒排索引外，Sorted文件夹内是raw sort结果，即只有一个ReduceTask
- output\_20210507\_13\_32内使用了随机采样和TotalOrderPartitioner均匀分配任务
- output\_20210507\_23\_09为了截图重新提交

## 唯一 & 穿着

- 唯一 808.78, 追风筝的人第五章第四节:146; 追风筝的人第三章第四节:134; 追风筝的人第十二章第六节:243; 追风筝的人第八章第五节:252; 追风筝的人第八章第三节:313; 追风筝的人第三章第二节:307; 追风筝的人第十二章第三节:262; 追风筝的人第十章第三节:328; 追风筝的人第七章第五节:314; 追风筝的人第二章第三节:166; 追风筝的人第七章第一节:300; 追风筝的人第十一章第三节:294; 追风筝的人第七章第三节:267; 追风筝的人第七章第四节:339; 追风筝的人第十一章第五节:236; 追风筝的人第四章第二节:288; 追风筝的人第五章第一节:275; 追风筝的人第三章第三节:252; 追风筝的人第七章第六节:265; 追风筝的人第五章第三节:335; 追风筝的人第二章第二节:338; 二零零一太空漫游第五部第三章:413; 茶花女第二十一章:701; 红与黑第四十五章:644; 红与黑第二十八章:656; 红与黑第一章:287; 追风筝的人第四章第三节:277; 茶花女第十二章:701; 二零零一太空漫游第三部第五章:459; 追风筝的人第十章第一节:293; 红与黑第四十四章:721; 茶花女第十章:807; 二零零一太空漫游第四部第八章:251; 二零零一太空漫游第三部第三章:494; 二零零一太空漫游第四部第六章:520; 追风筝的人第十二章第五节:259; 茶花女第七章:829; 二零零一太空漫游第四部第二章:474; 二零零一太空漫游第五部第二章:277; 二零零一太空漫游第六部第四章:506; 茶花女第二十五章:774; 红与黑第十章:372; 茶花女第十四章:845; 茶花女第十六章:725; 红与黑第二十七章:436; 红与黑第十七章:490; 红与黑第十九章:906; 茶花女第二十三章:722; 二零零一太空漫游第六

部第二章:495; 茶花女第十一章:862; 二零零一太空漫游第二部第五章:242; 茶花女第十三章:916; 红与黑第七章:1138; 钢铁是怎样炼成的第二部第九章:657; 二零零一太空漫游第二部第一章:582; 红与黑第三章:480; 二零零一太空漫游第二部第四章:591; 二零零一太空漫游第二部第六章:558; 追风筝的人第十一章第四节:245; 红与黑第二十六章:1056; 茶花女第二十四章:919; 红与黑第十四章:301; 二零零一太空漫游第三部第四章:287; 二零零一太空漫游第四部第七章:695; 二零零一太空漫游第六部第三章:719; 追风筝的人第六章第一节:292; 二零零一太空漫游第二部第三章:718; 红与黑第十八章:1347; 红与黑第十六章:465; 茶花女第二十六章:1217; 红与黑第二十二章:1438; 红与黑第二十三章:1469; 红与黑第二十一章:1446; 红与黑第四章:356; 红与黑第二十九章:1631; 钢铁是怎样炼成的第一部第五章:1774; 钢铁是怎样炼成的第一部第一章:1881; 钢铁是怎样炼成的第一部第二章:1798; 红与黑第二十章:477; 红与黑第三十七章:479; 钢铁是怎样炼成的第一部第四章:2043; 钢铁是怎样炼成的第二部第六章:2364; 钢铁是怎样炼成的第一部第九章:2382; 钢铁是怎样炼成的第二部第八章:2461; 二零零一太空漫游第二部第二章:306; 钢铁是怎样炼成的第二部第一章:2515; 钢铁是怎样炼成的第一部第八章:2640; 红与黑第十一章:463; 钢铁是怎样炼成的第一部第三章:2814; 钢铁是怎样炼成的第一部第七章:2915; 二零零一太空漫游第一部第五章:149; 钢铁是怎样炼成的第二部第七章:3197; 钢铁是怎样炼成的第二部第五章:3293; 钢铁是怎样炼成的第一部第六章:3648; 钢铁是怎样炼成的第二部第四章:4140; 红与黑第四十三章:450; 钢铁是怎样炼成的第二部第二章:4460; 钢铁是怎样炼成的第二部第三章:4723; 红与黑第三十章:21320; 红与黑第三十一章:448; 茶花女第一章:486; 追风筝的人第五章第二节:274; 红与黑第三十三章:499; 茶花女第八章:578; 茶花女第五章:59; 红与黑第十五章:380; 红与黑第三十五章:525; 茶花女第二章:563; 二零零一太空漫游第五部第一章:253; 茶花女第十九章:572; 红与黑第三十二章:521; 二零零一太空漫游第四部第五章:257; 茶花女第六章:687; 二零零一太空漫游第一部第三章:338; 追风筝的人第九章第一节:261; 红与黑第四十二章:524; 茶花女第三章:660; 追风筝的人第六章第二节:290; 二零零一太空漫游第一部第二章:361; 茶花女第十七章:655; 二零零一太空漫游第四部第一章:394; 茶花女第十五章:647; 二零零一太空漫游第五部第六章:143; 红与黑第二章:228; 红与黑第五章:788; 二零零一太空漫游第一部第四章:409; 二零零一太空漫游第三部第六章:232; 红与黑第十二章:709; 二零零一太空漫游第六部第一章:414; 追风筝的人第十章第二节:347; 茶花女第十八章:654; 二零零一太空漫游第一部第一章:432; 红与黑第三十八章:570; 红与黑第三十四章:662; 茶花女第四章:756; 红与黑第八章:730; 二零零一太空漫游第五部第五章:154; 红与黑第二十四章:668; 茶花女第二十二章:599; 追风筝的人第十二章第四节:290; 红与黑第六章:777; 红与黑第二十五章:684; 追风筝的人第七章第二节:282; 红与黑第九章:813; 红与黑第四十一章:656; 二零零一太空漫游第五部第四章:228; 茶花女第九章:796; 二零零一太空漫游第二部第七章:389; 红与黑第四十章:407; 红与黑第三十六章:699; 追风筝的人第十三章结局二:209; 二零零一太空漫游第三部第一章:490; 红与黑第十三章:477; 红与黑第三十九章:431; 追风筝的人第十一章第一节:272; 二零零一太空漫游第四部第四章:309; 追风筝的人第八章第一节:306; 追风筝的人第十一章第二节:269; 追风筝的人第八章第六节:211; 茶花女第二十章:503; 追风筝的人第十三章结局三:192; 二零零一太空漫游第五部第七章:274; 二零零一太空漫游第四部第三章:263; 追风筝的人第十二章第一节:225; 追风筝的人第八章第四节:327; 追风筝的人第二章第一节:276; 茶花女第二十七章:169; 追风筝的人第十二章第二节:235; 追风筝的人第十章第四节:280; 追风筝的人第十三章结局一:186; 追风筝的人第六章第三节:325; 追风筝的人第九章第二节:321; 追风筝的人第四章第一节:247; 追风筝的人第十三章结局四:100; 追风筝的人第八章第二节:299; 二零零一太空漫游第一部第六章:185; 追风筝的人第三章第一节:269; 二零零一太空漫游第三部第二章:260; 追风筝的人第一章:59; 追风筝的人第九章第三节:83; 追风筝的人第六章第四节:8; 追风筝的人第十二章第七节:83


- 穿着 2059.58, 追风筝的人第五章第四节:330; 追风筝的人第三章第四节:325; 追风筝的人第十二章第六节:720; 追风筝的人第八章第五节:747; 追风筝的人第八章第三节:839; 追风筝的人第三章第二节:779; 追风筝的人第十二章第三节:784; 追风筝的人第十章第三节:784; 追风筝的人第七章第五节:774; 追风筝的人第二章第三节:421; 追风筝的人第七章第一节:760; 追风筝的人第十一章第三节:830; 追风筝的人第七章第三节:769; 追风筝的人第七章第四节:800; 追风筝的人第十一章第五节:676; 追风筝的人第四章第二节:798; 追风筝的人第五章第一节:768; 追风筝的人第三章第三节:701; 追风筝的人第七章第六节:719; 追风筝的人第五章第三节:819; 追风筝的人第二章第二节:801; 二零零一太空漫游第五部第三章:1151; 茶花女第二十一章:1879; 红与黑第四十五章:1843; 红与黑第二十八章:1770; 红与黑第一章:810; 追风筝的人第四章第三节:769; 茶花女第十二章:2030; 二零零一太空漫游第三部第五章:1209; 追风筝的人第十章第一节:837; 红与黑第四十四章:1927; 茶

花女第十章:2429; 二零零一太空漫游第四部第八章:723; 二零零一太空漫游第三部第三章:1243; 二零零一太空漫游第四部第六章:1262; 追风筝的人第十二章第五节:792; 茶花女第七章:2435; 二零零一太空漫游第四部第二章:1269; 二零零一太空漫游第五部第二章:698; 二零零一太空漫游第六部第四章:1344; 茶花女第二十五章:2175; 红与黑第十章:840; 茶花女第十四章:2411; 茶花女第十六章:2326; 红与黑第二十七章:991; 红与黑第十七章:1188; 红与黑第十九章:2395; 茶花女第二十三章:2243; 二零零一太空漫游第六部第二章:1379; 茶花女第十一章:2439; 二零零一太空漫游第二部第五章:711; 茶花女第十三章:2563; 红与黑第七章:2903; 钢铁是怎样炼成的第二部第九章:1610; 二零零一太空漫游第二部第一章:1504; 红与黑第三章:1316; 二零零一太空漫游第二部第四章:1539; 二零零一太空漫游第二部第六章:1595; 追风筝的人第十一章第四节:768; 红与黑第二十六章:2599; 茶花女第二十四章:2783; 红与黑第十四章:799; 二零零一太空漫游第三部第四章:706; 二零零一太空漫游第四部第七章:1686; 二零零一太空漫游第六部第三章:1799; 追风筝的人第六章第一节:859; 二零零一太空漫游第二部第三章:1911; 红与黑第十八章:3254; 红与黑第十六章:1267; 茶花女第二十六章:3518; 红与黑第二十二章:3276; 红与黑第二十三章:3571; 红与黑第二十一章:3645; 红与黑第四章:932; 红与黑第二十九章:3928; 钢铁是怎样炼成的第一部第五章:4186; 钢铁是怎样炼成的第一部第一章:4359; 钢铁是怎样炼成的第一部第二章:4341; 红与黑第二十章:1182; 红与黑第三十七章:1151; 钢铁是怎样炼成的第一部第四章:5024; 钢铁是怎样炼成的第二部第六章:5381; 钢铁是怎样炼成的第一部第九章:5683; 钢铁是怎样炼成的第二部第八章:5793; 二零零一太空漫游第二部第二章:740; 钢铁是怎样炼成的第二部第一章:5911; 钢铁是怎样炼成的第一部第八章:6415; 红与黑第十一章:1326; 钢铁是怎样炼成的第一部第三章:6685; 钢铁是怎样炼成的第一部第七章:6998; 二零零一太空漫游第一部第五章:334; 钢铁是怎样炼成的第二部第七章:7375; 钢铁是怎样炼成的第二部第五章:7640; 钢铁是怎样炼成的第一部第六章:8469; 钢铁是怎样炼成的第二部第四章:9446; 红与黑第四十三章:1230; 钢铁是怎样炼成的第二部第二章:10338; 钢铁是怎样炼成的第二部第三章:11236; 红与黑第三十章:54337; 红与黑第三十一章:1282; 茶花女第一章:1510; 追风筝的人第五章第二节:782; 红与黑第三十三章:1321; 茶花女第八章:1628; 茶花女第五章:179; 红与黑第十五章:932; 红与黑第三十五章:1306; 茶花女第二章:1633; 二零零一太空漫游第五部第一章:611; 茶花女第十九章:1571; 红与黑第三十二章:1440; 二零零一太空漫游第四部第五章:589; 茶花女第六章:1734; 二零零一太空漫游第一部第三章:921; 追风筝的人第九章第一节:754; 红与黑第四十二章:1444; 茶花女第三章:1765; 追风筝的人第六章第二节:759; 二零零一太空漫游第一部第二章:930; 茶花女第十七章:1744; 二零零一太空漫游第四部第一章:983; 茶花女第十五章:1807; 二零零一太空漫游第五部第六章:369; 红与黑第二章:742; 红与黑第五章:1899; 二零零一太空漫游第一部第四章:1023; 二零零一太空漫游第三部第六章:650; 红与黑第十二章:1767; 二零零一太空漫游第六部第一章:1024; 追风筝的人第十章第二节:806; 茶花女第十八章:1862; 二零零一太空漫游第一部第一章:1076; 红与黑第三十八章:1675; 红与黑第三十四章:1740; 茶花女第四章:2123; 红与黑第八章:2050; 二零零一太空漫游第五部第五章:410; 红与黑第二十四章:1650; 茶花女第二十二章:1752; 追风筝的人第十二章第四节:748; 红与黑第六章:2053; 红与黑第二十五章:1651; 追风筝的人第七章第二节:774; 红与黑第九章:2109; 红与黑第四十一章:1742; 二零零一太空漫游第五部第四章:629; 茶花女第九章:2155; 二零零一太空漫游第二部第七章:1116; 红与黑第四十章:1048; 红与黑第三十六章:1788; 追风筝的人第十三章结局二:553; 二零零一太空漫游第三部第一章:1160; 红与黑第十三章:1329; 红与黑第三十九章:1225; 追风筝的人第十一章第一节:775; 二零零一太空漫游第四部第四章:759; 追风筝的人第八章第一节:753; 追风筝的人第十一章第二节:846; 追风筝的人第八章第六节:629; 茶花女第二十章:1419; 追风筝的人第十三章结局三:586; 二零零一太空漫游第五部第七章:664; 二零零一太空漫游第四部第三章:593; 追风筝的人第十二章第一节:710; 追风筝的人第八章第四节:911; 追风筝的人第二章第一节:831; 茶花女第二十七章:464; 追风筝的人第十二章第二节:774; 追风筝的人第十章第四节:812; 追风筝的人第十三章结局一:552; 追风筝的人第六章第三节:787; 追风筝的人第九章第二节:779; 追风筝的人第四章第一节:767; 追风筝的人第十三章结局四:278; 追风筝的人第八章第二节:802; 二零零一太空漫游第一部第六章:423; 追风筝的人第三章第一节:820; 二零零一太空漫游第三部第二章:686; 追风筝的人第一章:231; 追风筝的人第九章第三节:248; 追风筝的人第六章第四节:28; 追风筝的人第十二章第七节:220

## MapReduce 截图

- output 文件夹运行记录





Application application\_1604923458283\_1179

Logged in as: dr.who

Cluster

- About
- Nodes
- Node Labels
- Applications
- NEW
- NEW SAVING
- SUBMITTED
- ACCEPTED
- RUNNING
- FINISHED
- FAILED
- KILLED
- Scheduler

Tools

### Kill Application

Application Overview

User: nju\_st42  
Name: invert index  
Application Type: MAPREDUCE  
Application Tags:  
YarnApplicationState: FINISHED  
Queue: root.2021team42  
FinalStatus Reported by AM: SUCCEEDED  
Started: Wed May 05 13:05:10 +0800 2021  
Elapsed: 1mins, 19sec  
Tracking URL: History  
Diagnostics:

Application Metrics

Total Resource Preempted: <memory:0, vCores:0>  
Total Number of Non-AM Containers Preempted: 0  
Total Number of AM Containers Preempted: 0  
Resource Preempted from Current Attempt: <memory:0, vCores:0>  
Number of Non-AM Containers Preempted from Current Attempt: 0  
Aggregate Resource Allocation: 823299 MB-seconds, 625 vcore-seconds

Show 20 entries

Attempt ID

Started

Node

Logs

Blacklisted Nodes


appattempt_1604923458283_1179_000001	Wed May 5 13:05:10 +0800 2021	http://slave018:8042	Logs	N/A
--------------------------------------	-------------------------------	----------------------	------	-----

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

图2 Output Folder Run Record

- output\_20210507\_10\_56 - Sorted Record



Application application\_1604923458283\_1318

Logged in as: dr.who

Cluster

- About
- Nodes
- Node Labels
- Applications
- NEW
- NEW SAVING
- SUBMITTED
- ACCEPTED
- RUNNING
- FINISHED
- FAILED
- KILLED
- Scheduler

Tools

### Kill Application

Application Overview

User: nju\_st42  
Name: Sorted  
Application Type: MAPREDUCE  
Application Tags:  
YarnApplicationState: FINISHED  
Queue: root.2021team42  
FinalStatus Reported by AM: SUCCEEDED  
Started: Fri May 07 10:55:21 +0800 2021  
Elapsed: 19sec  
Tracking URL: History  
Diagnostics:

Application Metrics

Total Resource Preempted: <memory:0, vCores:0>  
Total Number of Non-AM Containers Preempted: 0  
Total Number of AM Containers Preempted: 0  
Resource Preempted from Current Attempt: <memory:0, vCores:0>  
Number of Non-AM Containers Preempted from Current Attempt: 0  
Aggregate Resource Allocation: 67226 MB-seconds, 38 vcore-seconds

Show 20 entries

Attempt ID

Started

Node

Logs

Blacklisted Nodes


appattempt_1604923458283_1318_000001	Fri May 7 10:55:21 +0800 2021	http://slave003:8042	Logs	N/A
--------------------------------------	-------------------------------	----------------------	------	-----

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

图3 output\_20210507\_10\_56 - Sorted Record

- Output\_20210507\_23\_09 Record



Application application\_1604923458283\_1438

Logged in as: dr.who

Cluster

- About
- Nodes
- Node Labels
- Applications
- NEW
- NEW SAVING
- SUBMITTED
- ACCEPTED
- RUNNING
- FINISHED
- FAILED
- KILLED
- Scheduler

Tools

### Kill Application

Application Overview

User: nju\_st42  
Name: Invert Index  
Application Type: MAPREDUCE  
Application Tags:  
YarnApplicationState: FINISHED  
Queue: root.2021team42  
FinalStatus Reported by AM: SUCCEEDED  
Started: Fri May 07 23:05:56 +0800 2021  
Elapsed: 1mins, 18sec  
Tracking URL: History  
Diagnostics:

Application Metrics

Total Resource Preempted: <memory:0, vCores:0>  
Total Number of Non-AM Containers Preempted: 0  
Total Number of AM Containers Preempted: 0  
Resource Preempted from Current Attempt: <memory:0, vCores:0>  
Number of Non-AM Containers Preempted from Current Attempt: 0  
Aggregate Resource Allocation: 822329 MB-seconds, 628 vcore-seconds

Show 20 entries

Attempt ID

Started

Node

Logs


Blacklisted Nodes

appattempt_1604923458283_1438_000001	Fri May 7 23:05:56 +0800 2021	http://slave013:8042	Logs	N/A
--------------------------------------	-------------------------------	----------------------	------	-----

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

图4.1 Output\_20210507\_23\_09\_InvertedIndex



Application application\_1604923458283\_1439

Logged in as: dr.who

Cluster

- About
- Nodes
- Node Labels
- Applications
- NEW
- NEW SAVING
- SUBMITTED
- ACCEPTED
- RUNNING
- FINISHED
- FAILED
- KILLED
- Scheduler

Tools

### Kill Application

Application Overview

User: nju\_st42  
Name: Convert Freq to Key  
Application Type: MAPREDUCE  
Application Tags:  
YarnApplicationState: FINISHED  
Queue: root.2021team42  
FinalStatus Reported by AM: SUCCEEDED  
Started: Fri May 07 23:07:16 +0800 2021  
Elapsed: 16sec  
Tracking URL: History  
Diagnostics:

### Application Metrics

Total Resource Preempted: <memory:0, vCores:0>  
Total Number of Non-AM Containers Preempted: 0  
Total Number of AM Containers Preempted: 0  
Resource Preempted from Current Attempt: <memory:0, vCores:0>  
Number of Non-AM Containers Preempted from Current Attempt: 0  
Aggregate Resource Allocation: 60783 MB-seconds, 35 vcore-seconds

Show 20 entries


Attempt ID
Started
Node
Logs
Blacklisted Nodes

appattempt_1604923458283_1439_000001	Fri May 7 23:07:16 +0800 2021	http://slave018:8042	Logs	N/A
--------------------------------------	-------------------------------	----------------------	------	-----

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

图4.2 Output\_20210507\_23\_09\_ConvertFreqToKey



Application application\_1604923458283\_1440

Logged in as: dr.who

Cluster

- About
- Nodes
- Node Labels
- Applications
- NEW
- NEW SAVING
- SUBMITTED
- ACCEPTED
- RUNNING
- FINISHED
- FAILED
- KILLED
- Scheduler

Tools

### Kill Application

Application Overview

User: nju\_st42  
Name: Sorted  
Application Type: MAPREDUCE  
Application Tags:  
YarnApplicationState: FINISHED  
Queue: root.2021team42  
FinalStatus Reported by AM: SUCCEEDED  
Started: Fri May 07 23:07:35 +0800 2021  
Elapsed: 16sec  
Tracking URL: History  
Diagnostics:

### Application Metrics

Total Resource Preempted: <memory:0, vCores:0>  
Total Number of Non-AM Containers Preempted: 0  
Total Number of AM Containers Preempted: 0  
Resource Preempted from Current Attempt: <memory:0, vCores:0>  
Number of Non-AM Containers Preempted from Current Attempt: 0  
Aggregate Resource Allocation: 67925 MB-seconds, 40 vcore-seconds

Show 20 entries

Attempt ID
Started
Node
Logs
Blacklisted Nodes

appattempt_1604923458283_1440_000001	Fri May 7 23:07:35 +0800 2021	http://slave007:8042	Logs	N/A
--------------------------------------	-------------------------------	----------------------	------	-----

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

图4.3 Output\_20210507\_23\_09\_Sorted

## 单级测试脚本

```
cd /home/hadoop/hadoop_workspace/homework2
rm InvertedIndex-1.0.0.jar
cp /mnt/d/Java_workspace/IdeaProjects/InvertedIndex-1.0.0/out/artifacts/InvertedIndex_1_0_0_jar/InvertedIndex-1.0.0.jar ./
hadoop dfs -rm -r test_out
hadoop jar InvertedIndex-1.0.0.jar MainClass test_in test_out
rm -r test_out
hadoop dfs -copyToLocal test_out ./test_out
```