1)We have a dataset of sales of different TV sets across different locations. Records look like: Samsung|Optima|14|Madhya Pradesh|132401|14200 The fields are arranged like: Company Name|Product Name|Size in inches|State|Pin Code|Price Sales\_TV:

Sales of different TV Task 4 : Write a Mapreduce program to modify Sales of different TV Task 2 ( refer session 5 assignment 2 ) to use a custom partitioner with 4 reducers. Make sure that all records whose company name starts with A-F (upper or lower case) should go to 1st reducer, those starting with G-L to 2nd reducer, those starting with M-R to 3rd reducer and others to 4th reducer. Sales of different TV Task 5 : Modify Sales of different TV Task 2 (refer session 5, assignment 2) to take advantage of Combiner.

|  |
| --- |
|  |
|  |
|  |  |
|  |  |
|  |  |
| //driver class |  |
|  | import org.apache.hadoop.fs.Path; |
|  | import org.apache.hadoop.conf.\*; |
|  | import org.apache.hadoop.mapreduce.Job; |
|  | import org.apache.hadoop.mapreduce.lib.input.FileInputFormat; |
|  | import org.apache.hadoop.mapreduce.lib.input.TextInputFormat; |
|  | import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat; |
|  | import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat; |
|  | import org.apache.hadoop.io.IntWritable; |
|  | import org.apache.hadoop.io.Text; |
|  |  |
|  | public class task7 { |
|  |  |
|  | public static void main(String[] args) throws Exception { |
|  | Configuration conf = new Configuration(); |
|  | Job job = new Job(conf, "Assignment"); |
|  | job.setJarByClass(task7.class); |
|  |  |
|  | job.setMapOutputKeyClass(Text.class); |
|  | job.setMapOutputValueClass(IntWritable.class); |
|  |  |
|  | job.setNumReduceTasks(4); |
|  | job.setPartitionerClass(part7.class); |
|  |  |
|  |  |
|  | job.setOutputKeyClass(Text.class); |
|  | job.setOutputValueClass(IntWritable.class); |
|  | job.setMapperClass(map7.class); |
|  | job.setReducerClass(reduce7.class); |
|  |  |
|  | job.setInputFormatClass(TextInputFormat.class); |
|  | job.setOutputFormatClass(TextOutputFormat.class); |
|  |  |
|  | FileInputFormat.addInputPath(job, new Path(args[0])); |
|  |  |
|  | Path outputPath = new Path(args[1]); |
|  | FileOutputFormat.setOutputPath(job, outputPath); |
|  | outputPath.getFileSystem(conf).delete(outputPath, true); |
|  |  |
|  |  |
|  |  |
|  | job.waitForCompletion(true); |
|  | } |
|  | } |
|  |  |
|  |  |
|  | //mapper class |
|  |  |
|  |  |
|  |  |
|  | import java.io.IOException; |
|  |  |
|  | import org.apache.hadoop.io.IntWritable; |
|  | import org.apache.hadoop.io.LongWritable; |
|  | import org.apache.hadoop.io.Text; |
|  | import org.apache.hadoop.mapreduce.\*; |
|  |  |
|  | public class map7 extends Mapper<LongWritable, Text, Text, IntWritable> { |
|  | Text outKey = new Text(); |
|  | IntWritable outValue = new IntWritable(); |
|  |  |
|  | public void map(LongWritable key, Text value, Context context) |
|  | throws IOException, InterruptedException { |
|  | String[] lineArray = value.toString().split("\\|"); |
|  | outKey.set(lineArray[0]); |
|  | outValue.set(1); |
|  | context.write(outKey, outValue); |
|  | } |
|  | } |
|  |  |
|  | //reducer class |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  | import java.io.IOException; |
|  |  |
|  | import org.apache.hadoop.io.IntWritable; |
|  | import org.apache.hadoop.io.Text; |
|  | import org.apache.hadoop.mapreduce.Reducer; |
|  |  |
|  | public class reduce7 extends Reducer<Text, IntWritable, Text, IntWritable> |
|  | { |
|  | IntWritable outValue = new IntWritable(); |
|  |  |
|  | public void reduce(Text key, Iterable<IntWritable> values,Context context) throws IOException, InterruptedException |
|  | { |
|  | int sum = 0; |
|  | for (IntWritable value : values) { |
|  | sum += value.get(); |
|  | } |
|  |  |
|  | outValue.set(sum); |
|  | context.write(key, outValue); |
|  | } |
|  | } |
|  |  |
|  |  |
|  |  |
|  |  |
|  | //partioner class |
|  |  |
|  |  |
|  |  |
|  | import org.apache.hadoop.io.IntWritable; |
|  | import org.apache.hadoop.io.Text; |
|  | import org.apache.hadoop.mapreduce.Partitioner; |
|  |  |
|  |  |
|  |  |
|  | public class part7 extends Partitioner<Text,IntWritable> { |
|  |  |
|  |  |
|  | public int getPartition(Text key, IntWritable value, int numReduceTasks) { |
|  |  |
|  | String name = key.toString(); |
|  | String firstChar = (Character.toString((name.charAt(0)))); |
|  |  |
|  |  |
|  |  |
|  | if (firstChar.matches("[A-F?i]")){ |
|  | return 0; |
|  | } else if (firstChar.matches("[G-L?i]")) { |
|  | return 1; |
|  | } else if (firstChar.matches("[M-R?i]")) { |
|  | return 2; |
|  | } else { |
|  | return 3; |
|  | } |
|  | } |
|  |  |
|  |  |
|  |  |
|  |  |
|  | } |
|  |  |
|  |  |
|  |  |
|  | //output |
|  |  |
|  | [acadgild@localhost Desktop]$ hadoop fs -cat /user/acadgild/7op1/part-r-00000 |
|  | Akai 1 |
|  | [acadgild@localhost Desktop]$ hadoop fs -cat /user/acadgild/7op1/part-r-00001 |
|  | Lava 3 |
|  | [acadgild@localhost Desktop]$ hadoop fs -cat /user/acadgild/7op1/part-r-00002 |
|  | NA 1 |
|  | Onida 4 |
|  | [acadgild@localhost Desktop]$ hadoop fs -cat /user/acadgild/7op1/part-r-00003 |
|  | Samsung 7 |
|  | Zen 2   |  | | --- | |  | | //driver class | |  | import org.apache.hadoop.fs.Path; | |  | import org.apache.hadoop.conf.\*; | |  | import org.apache.hadoop.mapreduce.Job; | |  | import org.apache.hadoop.mapreduce.lib.input.FileInputFormat; | |  | import org.apache.hadoop.mapreduce.lib.input.TextInputFormat; | |  | import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat; | |  | import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat; | |  | import org.apache.hadoop.io.IntWritable; | |  | import org.apache.hadoop.io.Text; | |  |  | |  | public class task7 { | |  |  | |  | public static void main(String[] args) throws Exception { | |  | Configuration conf = new Configuration(); | |  | Job job = new Job(conf, "DemoTask1"); | |  | job.setJarByClass(task7.class); | |  |  | |  |  | |  | job.setMapOutputKeyClass(Text.class); | |  | job.setMapOutputValueClass(IntWritable.class); | |  |  | |  | job.setOutputKeyClass(Text.class); | |  | job.setOutputValueClass(IntWritable.class); | |  | job.setMapperClass(map7.class); | |  | job.setCombinerClass(reduce7.class);// combiner class set | |  | job.setReducerClass(reduce7.class); | |  | job.setInputFormatClass(TextInputFormat.class); | |  | job.setOutputFormatClass(TextOutputFormat.class); | |  |  | |  | FileInputFormat.addInputPath(job, new Path(args[0])); | |  | FileOutputFormat.setOutputPath(job,new Path(args[1])); | |  |  | |  | /\* | |  | Path out=new Path(args[1]); | |  | out.getFileSystem(conf).delete(out); | |  | \*/ | |  |  | |  | job.waitForCompletion(true);// wait for the job to end | |  |  | |  |  | |  |  | |  |  | |  | } | |  | } | |  |  | |  | //map | |  |  | |  | import java.io.IOException; | |  |  | |  | import org.apache.hadoop.io.IntWritable; | |  | import org.apache.hadoop.io.LongWritable; | |  | import org.apache.hadoop.io.Text; | |  | import org.apache.hadoop.mapred.TestJobCounters.NewSummer; | |  | import org.apache.hadoop.mapreduce.\*; | |  |  | |  | public class map7 extends Mapper<LongWritable, Text, Text, IntWritable> { | |  | IntWritable no1 = new IntWritable(1); | |  | public void map(LongWritable key, Text value, Context context) | |  | throws IOException, InterruptedException { | |  | String[] lineArray = value.toString().split("\\|"); | |  | String s = lineArray[0];//stroing the splitted values in strings | |  | String s1 = lineArray[1]; | |  | if(!(s.equals("NA") || s1.equals("NA")))//NA condition implemented | |  | { | |  | context.write(new Text(s), no1); | |  | } | |  | } | |  | } | |  |  | |  |  | |  | //reduce | |  |  | |  | import java.io.IOException; | |  | import org.apache.hadoop.io.Text; | |  | import org.apache.hadoop.io.IntWritable; | |  | import org.apache.hadoop.mapreduce.Reducer; | |  | public class reduce7 extends Reducer<Text,IntWritable,Text, IntWritable> { | |  | public void reduce(Text key, Iterable<IntWritable> value , Context con) throws IOException, InterruptedException | |  | { | |  | int sum =0; | |  | for(IntWritable v : value) | |  | { | |  | sum = sum + v.get(); | |  | } | |  | con.write(key, new IntWritable(sum)); | |  | } | |  |  | |  | } | |  |  | |  | //output | |  |  | |  | hadoop fs -cat /user/acadgild/7bop1/part-r-00000 | |  | Akai 1 | |  | Lava 3 | |  | Onida 3 | |  | Samsung 7 | |  | Zen 2 | |  |  | |
|  |  |

|  |  |
| --- | --- |
|  | Samsung|Optima|14|Madhya Pradesh|132401|14200 |
|  | Onida|Lucid|18|Uttar Pradesh|232401|16200 |
|  | Akai|Decent|16|Kerala|922401|12200 |
|  | Lava|Attention|20|Assam|454601|24200 |
|  | Zen|Super|14|Maharashtra|619082|9200 |
|  | Samsung|Optima|14|Madhya Pradesh|132401|14200 |
|  | Onida|Lucid|18|Uttar Pradesh|232401|16200 |
|  | Onida|Decent|14|Uttar Pradesh|232401|16200 |
|  | Onida|NA|16|Kerala|922401|12200 |
|  | Lava|Attention|20|Assam|454601|24200 |
|  | Zen|Super|14|Maharashtra|619082|9200 |
|  | Samsung|Optima|14|Madhya Pradesh|132401|14200 |
|  | NA|Lucid|18|Uttar Pradesh|232401|16200 |
|  | Samsung|Decent|16|Kerala|922401|12200 |
|  | Lava|Attention|20|Assam|454601|24200 |
|  | Samsung|Super|14|Maharashtra|619082|9200 |
|  | Samsung|Super|14|Maharashtra|619082|9200 |
|  | Samsung|Super|14|Maharashtra|619082|9200 |