|  |
| --- |
| import java.io.IOException; |
|  |  |
|  | import org.apache.hadoop.conf.Configuration; |
|  | import org.apache.hadoop.fs.Path; |
|  | import org.apache.hadoop.io.LongWritable; |
|  | import org.apache.hadoop.io.Text; |
|  | import org.apache.hadoop.mapreduce.Job; |
|  | import org.apache.hadoop.mapreduce.Mapper; |
|  | import org.apache.hadoop.mapreduce.lib.input.FileInputFormat; |
|  | import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat; |
|  | import org.apache.hadoop.mapreduce.lib.output.SequenceFileOutputFormat; |
|  |  |
|  | public class TextToSequence { |
|  | public static class SequenceMapper extends |
|  | Mapper<LongWritable, Text, LongWritable, Text> { |
|  | public void map(LongWritable key, Text value, Context context) |
|  | throws IOException, InterruptedException { |
|  | String record = value.toString(); |
|  | String[] parts = record.split(","); |
|  | int mykey = Integer.parseInt(parts[0]); |
|  | context.write(new LongWritable(mykey), new Text(parts[1])); |
|  | } |
|  | } |
|  |  |
|  | public static void main(String[] args) throws Exception { |
|  |  |
|  | Configuration conf = new Configuration(); |
|  | Job job = Job.getInstance(conf, "Convert Text to Sequence"); |
|  | job.setJarByClass(TextToSequence.class); |
|  | job.setMapperClass(SequenceMapper.class); |
|  | job.setNumReduceTasks(0); |
|  | job.setOutputKeyClass(LongWritable.class); |
|  | job.setOutputValueClass(Text.class); |
|  | job.setOutputFormatClass(SequenceFileOutputFormat.class); |
|  | FileInputFormat.addInputPath(job, new Path(args[0])); |
|  | FileOutputFormat.setOutputPath(job, new Path(args[1])); |
|  | System.exit(job.waitForCompletion(true) ? 0 : 1); |
|  | } |

}