# Java操作Linux命令分割合并文本文件及其他

## 1、说明

有时候会对很大的文本文件进行处理，用流一次性读入肯定是不可能的，内存吃不住，所以需要对文件进行分割、处理、合并，以下是写的工具类。

## 2、磁盘空间使用率获取

在处理文本之前，一定要备份一下，在备份之前要判断一下磁盘空间是否足够，用到的Linux命令是“df -hl -P”。命令详情请自己查阅，以下是Java代码：

|  |
| --- |
| **public** **static** **int** getDiskUsage (String filePath) **throws** Exception {  **int** diskUsage = 0;  ProcessBuilder builder = **null**;  Process ps = **null**;  **try** {  String[] cmd = { "/bin/sh", "-c", "df -hl -P" };  builder = **new** ProcessBuilder(cmd);  builder.redirectErrorStream(**true**);  ps = builder.start();  BufferedReader br = **new** BufferedReader(**new** InputStreamReader(ps.getInputStream(), "utf-8"));  **while** (**true**) {  String result = br.readLine();  **if** (result == **null**) {  **break**;  }  String[] r = result.split("\n");  **for** (**int** i = 0; i < r.length; i++) {  **if** (r[i].startsWith("/")) {  String[] t = r[i].split(" ");  **for** (**int** j = 0; j < t.length; j++) {  **if** (t[j].endsWith(filePath)) {  **if** (t[j - 1].contains("%")) {  diskUsage = Integer.*parseInt*(t[j - 1].replace("%", ""));  }  }  }  }  }  }  ps.waitFor();  **return** diskUsage;  } **catch** (Exception e) {  e.printStackTrace();  **throw** e;  } **finally** {  **if** (ps != **null**) {  ps.destroy();  }  }  } |

## 3、获取文件行数

获取总行数主要是为了能确认按行分割文件的话，能分割几个文件。使用的Linux命令是：find /home/leo -name "java.txt"|xargs cat|wc -l

Java代码如下：

|  |
| --- |
| **public** **static** **long** getLineNum(String filePath, String fileName) **throws** Exception {  **long** lineNums = 0l;  ProcessBuilder builder = **null**;  Process ps = **null**;  **try** {  String[] cmd = { "/bin/sh", "-c", "find " + filePath + " -name \"" + fileName + "\"|xargs cat|wc -l" };  builder = **new** ProcessBuilder(cmd);  builder.redirectErrorStream(**true**);  ps = builder.start();  BufferedReader stdoutReader = **new** BufferedReader(**new** InputStreamReader(ps.getInputStream(), "utf-8"));// linux终端的编码为utf-8  **while** (**true**) {  String outLine = stdoutReader.readLine();  **if** (outLine == **null**) {  **break**;  }  **if** (outLine.contains("No such file or directory")) {  **throw** **new** Exception("查询文件行数失败,文件不存在!");  } **else** {  lineNums = Long.*parseLong*(outLine);  }  }  ps.waitFor();  **return** lineNums;  } **catch** (Exception e) {  e.printStackTrace();  **throw** e;  } **finally** {  **if** (ps != **null**) {  ps.destroy();  }  }  } |

## 4、分割文件

这里使用了分割、改名两条命令，所以写成了一个sh，Java代码如下：

|  |
| --- |
| **public** **static** String splitFile(String filePath, String prefix) **throws** Exception {  String restr = "fail";  ProcessBuilder builder = **null**;  Process ps = **null**;  **try** {  String[] cmd = { "/bin/sh", "-c", "/home/leo/splitfile.sh " + filePath + " " + prefix };  builder = **new** ProcessBuilder(cmd);  builder.redirectErrorStream(**true**);  ps = builder.start();  BufferedReader stdoutReader = **new** BufferedReader(**new** InputStreamReader(ps.getInputStream(), "utf-8"));// linux终端的编码为utf-8  **while** (**true**) {  String outLine = stdoutReader.readLine();  **if** (outLine == **null**) {  **break**;  }  **if** (outLine.contains("No such file or directory")) {  **throw** **new** Exception("查询文件行数失败,文件不存在!");  } **else** {  restr = outLine;  }  }  ps.waitFor();  **return** restr;  } **catch** (Exception e) {  e.printStackTrace();  **throw** e;  } **finally** {  **if** (ps != **null**) {  ps.destroy();  }  }  } |

sh文件代码如下：

|  |
| --- |
| #!/bin/sh  #要分割的文件  filename=$1  #分割后文件的前缀  sdpre=$2  #每个文件分割4W行  split -l 40000 ${filename} -d -a 3 ${sdpre}  #给分割后的文件加扩展名  ls|grep ${sdpre}|xargs -n1 -i{} mv {} {}.txt  echo "done" |

注意不要忘了给sh赋权。

## 5、合并文件

使用的Linux命令是：cat /home/leo/sd\_000.txt /home/leo/sd\_001.txt /home/leo/sd\_002.txt /home/leo/sd\_003.txt /home/leo/sd\_004.txt > /home/leo/java2.txt

Java代码如下：

|  |
| --- |
| **public** **static** String catFiles(String fileList, String destFileName) **throws** Exception {  String restr = "done";  ProcessBuilder builder = **null**;  Process ps = **null**;  **try** {  // cat split00.txt split01.txt > split.txt  String[] cmd = { "/bin/sh", "-c", "cat " + fileList + " > " + destFileName };  builder = **new** ProcessBuilder(cmd);  builder.redirectErrorStream(**true**);  ps = builder.start();  BufferedReader stdoutReader = **new** BufferedReader(**new** InputStreamReader(ps.getInputStream(), "utf-8"));// linux终端的编码为utf-8  **while** (**true**) {  String outLine = stdoutReader.readLine();  **if** (outLine == **null**) {  **break**;  }  **if** (outLine.contains("No such file or directory")) {  **throw** **new** Exception("查询文件行数失败,文件不存在!");  } **else** {  restr = outLine;  }  }  ps.waitFor();  **return** restr;  } **catch** (Exception e) {  e.printStackTrace();  **throw** e;  } **finally** {  **if** (ps != **null**) {  ps.destroy();  }  }  } |

复制备份文件的代码就不写了。