字符处理

## 使用grep搜索文本

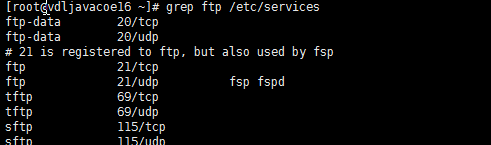
语法：grep [指定字串] [源文件]

功能描述：在文件中搜寻字串匹配的行并输出

常用选项：

* -c：统计包含匹配的行数
* -i：忽略大小写
* -v：反向匹配
* -w：匹配单词
* --color显示颜色

范例：# grep ftp /etc/services



## 使用sort排序

命令格式：sort 文件名 //对所有行进行排序

选项与参数：

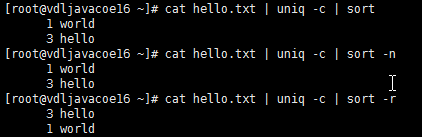
*  -f:忽略大小写
*  -n:数值类型进行排序
*  -r：反向排序
*  -t：指定分隔符（默认是制表符）
*  -k n[,m]:按照指定的字段范围排序。从n字段开始，m字段结束（默认到行尾）

## 使用uniq删除重复内容

命令格式：uniq 选项 文件名

选项：

* -c行首显示字符出现次数
* -i忽略大小写
* -d只显示重复行
* -u只显示不重复行

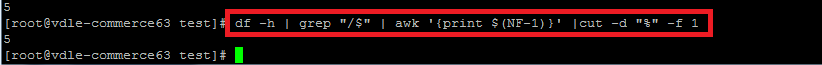


## 使用cut截取文本

cut命令格式：cut 选项 文件名

* -f 列号,列号：提取第几列
* -d 分隔符：按照指定分隔符分割列（默认的是\t）

例如：df -h | grep "/$" | awk '{print $(NF-1)}' |cut -d "%" -f 1



注意：要用空格来分隔的时候就不适合，且只能分割单个字符。

## 使用tr替换文本

tr用来从标准输入中通过替换或删除操作进行字符转换。tr主要用于删除文件中控制字符或进行字符转换。使用tr时要转换两个字符串：字符串1用于查询，字符串2用于处理各种转换。tr刚执行时，字符串1中的字符被映射到字符串2中的字符，然后转换操作开始。

tr命令格式为：

tr -c -d -s ["string1\_to\_translate\_from"] ["string2\_to\_translate\_to"] < input-file

选项：

* -c 用字符串1中字符集的补集替换此字符集，要求字符集为ASCII。
* -d 删除字符串1中所有输入字符。
* -s 删除所有重复出现字符序列，只保留第一个；即将重复出现字符串压缩为一个字符串。
* input-file是转换文件名。

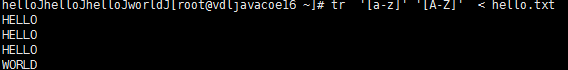
字符范围指定字符串1或字符串2的内容时，只能使用单字符或字符串范围或列表：

* [a-z] a-z内的字符组成的字符串。
* [A-Z] A-Z内的字符组成的字符串。
* [0-9] 数字串。
* \octal 一个三位的八进制数，对应有效的ASCII字符。
* [O\*n] 表示字符O重复出现指定次数n。因此[O\*2]匹配OO的字符串。
* tr中特定控制字符的不同表达方式

速记符含义八进制方式

* \a Ctrl-G 铃声\007
* \b Ctrl-H 退格符\010
* \f Ctrl-L 走行换页\014
* \n Ctrl-J 新行\012
* \r Ctrl-M 回车\015
* \t Ctrl-I tab键\011
* \v Ctrl-X \030

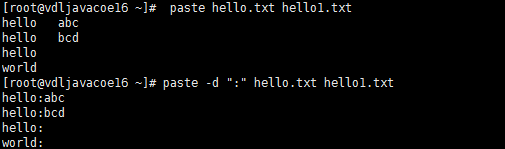
|  |
| --- |
| 例如：  #删除空行  cat file | tr -s "\n" > new\_file  #删除文件file中出现的"Snail"字符，则 file文件中出现的'S','n','a','i','l'字符都会被删除  cat file | tr -d "Snail" > new\_file  #将文件file中出现的"abc"替换为"xyz"  cat file | tr "abc" "xyz" > new\_file |



## 使用paste做文本合并

Paste的作用在于将文件按照行进行合并，使用tab隔开。可以通过-d来制定分隔符。例如：

|  |
| --- |
| paste hello.txt hello1.txt  paste -d ":" hello.txt hello1.txt |



## 使用split分割大文件

Linux可以通过split命令对文件进行分割，分割分为两种方式，一种是按行，适用于非二进制的文件；另一种是按文件的大小，适用于二进制文件。

|  |
| --- |
| #-l指定每个文件多少行，-d指定使用数字作为后缀  split -l 20000 catalina.out -d catalina.out\_  #-b指定每个文件的大小  split -b 64m catalina.out -d catalina.out\_ |

