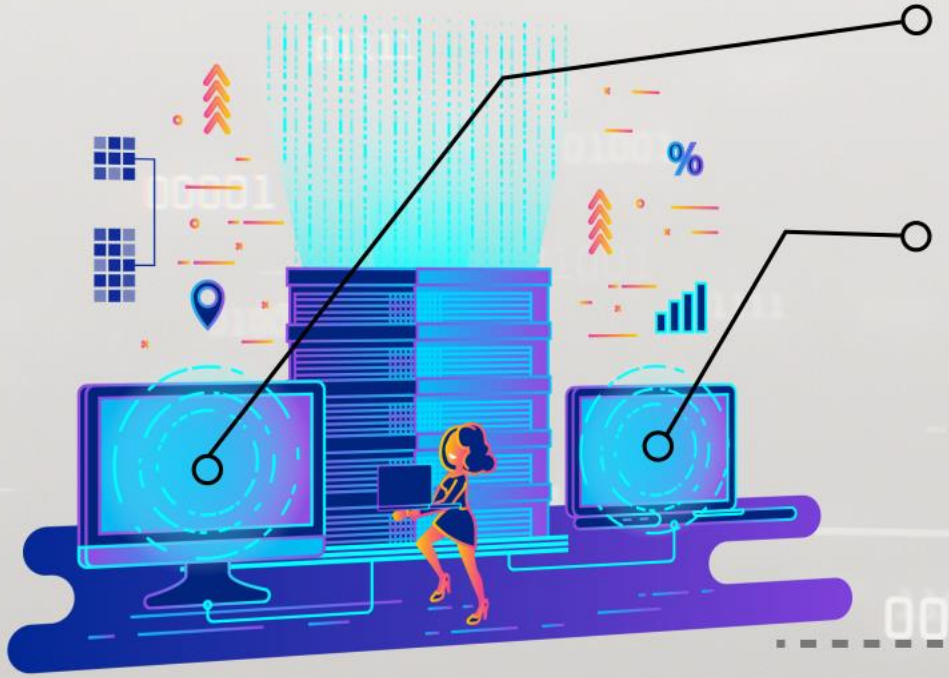
北京郵電大學
Beijing University of Posts and Telecommunications



文本数据的处理

蒋砚军 北京邮电大学计算机学院

tee：三通

- 功能
  - 将从标准输入stdin得到的数据抄送到标准输出stdout显示，同时存入磁盘文件中
- 应用举例
  - ./myap | tee myap.log

wc：字计数（word count）

功能

- 列出文件中一共有多少行，有多少个单词，多少字符
- 当指定的文件数大于1时，最后还列出一个合计
- 常用选项-l：只列出行计数

举例

- wc sum.c　　（1个文件）
- wc x.c makefile stat.sh （多个文件）
- wc -l *.c makefile start.sh
- ps -ef | wc -l （0个）
- ps -ef | grep liang | wc -l （0个）
- who | wc -l （0个）

sort：对文件内容排序

sort选项

- -n选项(Numberic):对于数字按照算术值大小排序，而不是按照字符串比较规则，例如123与67
- 可以选择行中某一部分作为排序关键字
- 选择升序或降序
- 字符串比较时对字母是否区分大小写
- 内排序外排序等算法参数选择（当数据量较大时，性能调优）

举例

- sort telno > telno1
- ls -s | sort | tail –10
- ls -s | sort -n | tail –10

tr：翻译字符

- **用法**

  `tr` *string1 string2*

  把标准输入拷贝到标准输出，*string1*中出现的字符替换为*string2*中的对应字符

- **例**

```
cat telnos | tr UVX uvx
```

- **例：用［］指定一个集合**

```
cat report | tr '[a-z]' '[A-Z]'
```

将小写字母改为大写字母

- **例：用\加三个八进制数字(类似C语言)表示一字符**

```
cat file1 | tr % '\012'
```

将%改为换行符，注意不要漏掉必需的单引号

uniq：筛选文件中的重复行

- 用法

  uniq *options*

  uniq *options input-file*

  uniq *options input-file output-file*

- 重复的行：紧邻的两行内容相同

- 选项

  -u （uniqe）只保留没有重复的行

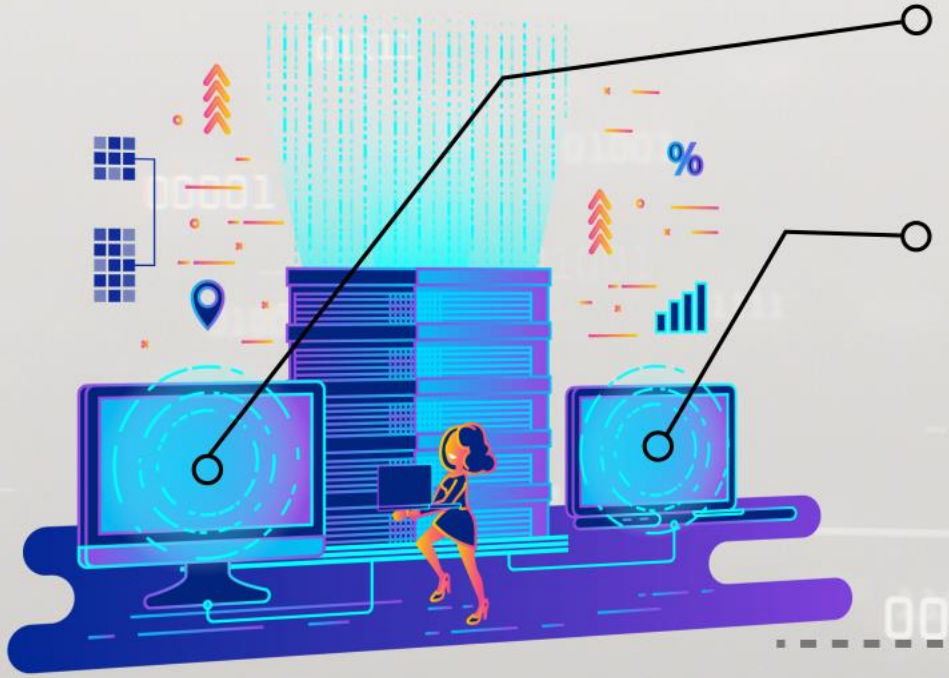
  -d （duplicated）只保留有重复的行（但只打印一次）

  没有以上两个选项，打印没有重复的行和有重复的行（但只打印一次)

  -c （count）计数同样的行出现几次

Linux
Windows
Windows
Linux
Linux
Linux
AIX

北京郵電大學
Beijing University of Posts and Telecommunications



谢 谢