

08 【实操篇-压缩和解压类】

1.打包（归档）和压缩

归档，也称为打包，指的是一个文件或目录的集合，而这个集合被存储在一个文件中。归档文件没有经过压缩，因此，它占用的空间是其中所有文件和目录的总和。

通常，归档总是会与系统（数据）备份联系在一起，不过，有关数据备份的内容，留到后续章节讲，本章仅学习归档命令的基本使用。

和归档文件类似，压缩文件也是一个文件和目录的集合，且这个集合也被存储在一个文件中，但它们的区别之处在于，压缩文件采用了不同的存储方式，使其所占用的磁盘空间比集合中所有文件大小的总和要小。

压缩是指利用算法将文件进行处理，已达到保留最大文件信息，而让文件体积变小的目的。其基本原理为，通过查找文件内的重复字节，建立一个相同字节的词典文件，并用一个代码表示。比如说，在压缩文件中，有不只一处出现了"C语言中文网"，那么，在压缩文件时，这个词就会用一个代码表示并写入词典文件，这样就可以实现缩小文件体积的目的。

由于计算机处理的信息是以二进制的形式表示的，因此，压缩软件就是把二进制信息中相同的字符串以特殊字符标记，只要通过合理的数学计算，文件的体积就能够被大大压缩。把一个或者多个文件用压缩软件进行压缩，形成一个文件压缩包，既可以节省存储空间，有方便在网络上传送。

如果你能够理解文件压缩的基本原理，那么很容易就能想到，对文件进行压缩，很可能损坏文件中的内容，因此，压缩又可以分为有损压缩和无损压缩。无损压缩很好理解，指的是压缩数据必须准确无误；有损压缩指的是即便丢失个别的数据，对文件也不会造成太大的影响。有损压缩广泛应用于动画、声音和图像文件中，典型代表就是影碟文件格式 mpeg、音乐文件格式 mp3 以及图像文件格式 jpg。

采用压缩工具对文件进行压缩，生成的文件称为压缩包，该文件的体积通常只有原文件的一半甚至更小。需要注意的是，压缩包中的数据无法直接使用，使用前需要利用压缩工具将文件数据还原，此过程又称解压缩。

Linux 下，常用归档命令有 2 个，分别是 tar 和 dd（相对而言，tar 的使用更为广泛）；常用的压缩命令有很多，比如 gzip、zip、bzip2 等。这些命令的详细用法，后续文件会做一一介绍。

注意，tar 命令也可以作为压缩命令，也很常用。

2.gzip/gunzip 压缩

2.1 gzip压缩文件

gzip 是 Linux 系统中经常用来对文件进行压缩和解压缩的命令，通过此命令压缩得到的新文件，其扩展名通常标记为".gz"。

再强调一下，gzip 命令只能用来压缩文件，不能压缩目录，即便指定了目录，也只能压缩目录内的所有文件。

gzip 命令的基本格式如下：

```
[root@localhost ~]# gzip [选项] 源文件
```

命令中的源文件，当进行压缩操作时，指的是普通文件；当进行解压缩操作时，指的是压缩文件。

选项	含义
-c	将压缩数据输出到标准输出中，并保留源文件。
-d	对压缩文件进行解压缩。
-r	递归压缩指定目录下以及子目录下的所有文件。
-v	对于每个压缩和解压缩的文件，显示相应的文件名和压缩比。
-l	对每一个压缩文件，显示以下字段：压缩文件的大小；未压缩文件的大小；压缩比；未压缩文件的名称。
-数字	用于指定压缩等级，-1 压缩等级最低，压缩比最差；-9 压缩比最高。默认压缩比是 -6。

【例 1】基本压缩。

gzip 压缩命令非常简单，甚至不需要指定压缩之后的压缩包名，只需指定源文件名即可。我们来试试：

```
[root@localhost ~]# gzip install.log
#压缩install.log 文件
[root@localhost ~]# ls
anaconda-ks.cfg install.log.gz install.log.syslog
#压缩文件生成，但是源文件也消失了
```

【例 2】保留源文件压缩。

在使用 gzip 命令压缩文件时，源文件会消失，从而生成压缩文件。能不能在压缩文件的时候，不让源文件消失？

```
[root@localhost ~]# gzip -c anaconda-ks.cfg >anaconda-ks.cfg.gz
#使用-c选项，但是不让压缩数据输出到屏幕上，而是重定向到压缩文件中，这样可以缩文件的同时不删除源文件
[root@localhost ~]# ls
anaconda-ks.cfg anaconda-ks.cfg.gz install.log.gz install.log.syslog
#可以看到压缩文件和源文件都存在
```

【例 3】压缩目录。

我们可能会想当然地认为 gzip 命令可以压缩目录。我们来试试：

```
[root@localhost ~]# mkdir test
[root@localhost ~]# touch test/test1
[root@localhost ~]# touch test/test2
[root@localhost ~]# touch test/test3 #建立测试目录，并在里面建立几个测试文件
[root@localhost ~]# gzip -r test/
#压缩目录，并没有报错
[root@localhost ~]# ls
anaconda-ks.cfg anaconda-ks.cfg.gz install.log.gz install.log.syslog test
#但是查看发现test目录依然存在，并没有变为压缩文件
[root@localhost ~]# ls test/
test1 .gz test2.gz test3.gz
#原来gzip命令不会打包目录，而是把目录下所有的子文件分别压缩
```

在 Linux 中，打包和压缩是分开处理的。而 gzip 命令只会压缩，不能打包，所以才会出现没有打包目录，而只把目录下的文件进行压缩的情况。

2.2 gunzip解压缩文件

gunzip 是一个使用广泛的解压缩命令，它用于解压被 gzip 压缩过的文件（扩展名为 .gz）。

对于解压被 gzip 压缩过的文件，还可以使用 gzip 自己，即 gzip -d 压缩包。

gunzip 命令的基本格式为：

```
[root@localhost ~]# gunzip [选项] 文件
```

选项	含义
-r	递归处理，解压缩指定目录下以及子目录下的所有文件。
-c	把解压缩后的文件输出到标准输出设备。
-f	强制解压缩文件，不理睬文件是否已存在等情况。
-l	列出压缩文件内容。
-v	显示命令执行过程。
-t	测试压缩文件是否正常，但不对其做解压缩操作。

【例 1】直接解压缩文件。

```
[root@localhost ~]# gunzip install.log.gz
```

当然，"gunzip -r"依然只会解压缩目录下的文件，而不会解打包。要想解压缩".gz"格式，还可以使用"gzip -d"命令，例如：

```
[root@localhost ~]# gzip -d anaconda-ks.cfg.gz
```

【例 2】要解压缩目录下的内容，则需使用 "-r" 选项，例如：

```
[root@localhost ~]# gunzip -r test/
```

注意，如果我们压缩的是一个纯文本文件，则可以直接使用 zcat 命令在不解压缩的情况下查看这个文本文件中的内容。例如：

```
[root@localhost ~]# zcat anaconda-ks.cfg.gz
```

总结

- (1) 只能压缩文件不能压缩目录
- (2) 不保留原来的文件
- (3) 同时多个文件会产生多个压缩包

3.zip/unzip 压缩

3.1 zip压缩文件或目录

我们经常会在 Windows 系统上使用“.zip”格式压缩文件，其实“.zip”格式文件是 Windows 和 Linux 系统都通用的压缩文件类型，属于几种主流的压缩格式（zip、rar等）之一，是一种相当简单的分别压缩每个文件的存储格式，

本节要讲的 zip 命令，类似于 Windows 系统中的 winzip 压缩程序，其基本格式如下：

```
[root@localhost ~]#zip [选项] 压缩包名 源文件或源目录列表
```

注意，zip 压缩命令需要手工指定压缩之后的压缩包名，注意写清楚扩展名，以便解压缩时使用。

选项	含义
-r	递归压缩目录，及将制定目录下的所有文件以及子目录全部压缩。
-m	将文件压缩之后，删除原始文件，相当于把文件移到压缩文件中。
-v	显示详细的压缩过程信息。
-q	在压缩的时候不显示命令的执行过程。
-压缩级别	压缩级别是从 1~9 的数字，-1 代表压缩速度更快，-9 代表压缩效果更好。
-u	更新压缩文件，即往压缩文件中添加新文件。

下面给大家举几个例子。

【例 1】zip 命令的基本使用。

```
[root@localhost ~]# zip ana.zip anaconda-ks.cfg
adding: anaconda-ks.cfg (deflated 37%)
#压缩
[root@localhost ~]# ll ana.zip
-rw-r--r-- 1 root root 935 6月 17:16:00 ana.zip
#压缩文件生成
```

不仅如此，所有的压缩命令都可以同时压缩多个文件，例如：

```
[root@localhost ~]# zip test.zip install.log install.log.syslog
adding: install.log (deflated 72%)
adding: install.log.syslog (deflated 85%)
#同时压缩多个文件到test.zip压缩包中
[root@localhost ~]#ll test.zip
-rw-r--r-- 1 root root 8368 6月 17:16:03 test.zip
#压缩文件生成
```

【例 2】使用 zip 命令压缩目录，需要使用“-r”选项，例如：

```
[root@localhost ~]# mkdir dir1
#建立测试目录
[root@localhost ~]# zip -r dir1.zip dir1
adding: dir1/(stored 0%)
#压缩目录
[root@localhost ~]# ls -dl dir1.zip
-rw-r--r-- 1 root root 160 6月 17:16:22 dir1.zip
#压缩文件生成
```

3.2 unzip解压zip文件

unzip 命令可以查看和解压缩 zip 文件。该命令的基本格式如下：

```
[root@localhost ~]# unzip [选项] 压缩包名
```

选项	含义
-d 目录名	将压缩文件解压到指定目录下。
-n	解压时并不覆盖已经存在的文件。
-o	解压时覆盖已经存在的文件，并且无需用户确认。
-v	查看压缩文件的详细信息，包括压缩文件中包含的文件大小、文件名以及压缩比等，但并不做解压操作。
-t	测试压缩文件有无损坏，但并不解压。
-x 文件列表	解压文件，但不包含文件列表中指定的文件。

【例 1】不论是文件压缩包，还是目录压缩包，都可以直接解压缩，例如：

```
[root@localhost ~]# unzip dir1.zip
Archive: dir1.zip
creating: dir1/
#解压缩
```

【例 2】使用 -d 选项手动指定解压缩位置，例如：

```
[root@localhost ~]# unzip -d /tmp/ ana.zip
Archive: ana.zip
inflating: /tmp/anaconda-ks.cfg
#把压缩包解压到指定位置
```

总结

zip 压缩命令在windows/linux都通用，可以压缩目录且保留源文件。

4.tar 打包

Linux 系统中，最常用的归档（打包）命令就是 tar，该命令可以将许多文件一起保存到一个单独的磁带或磁盘中进行归档。不仅如此，该命令还可以从归档文件中还原所需文件，也就是打包的反过程，称为解打包。

使用 tar 命令归档的包通常称为 tar 包（tar 包文件都是以“.tar”结尾的）。

4.1 tar命令做打包操作

当 tar 命令用于打包操作时，该命令的基本格式为：

```
[root@localhost ~]#tar [选项] 源文件或目录
```

选项	含义
-z	打包同时压缩
-c	将多个文件或目录进行打包。
-A	追加 tar 文件到归档文件。
-f 包名	指定包的文件名。包的扩展名是用来给管理员识别格式的，所以一定要正确指定扩展名；
-v	显示打包文件过程；

需要注意的是，在使用 tar 命令指定选项时可以不在选项前面输入“-”。例如，使用“cvf”选项和“-cvf”起到的作用一样。

下面给大家举几个例子，一起来看看如何使用 tar 命令打包文件和目录。

【例 1】打包文件和目录。

```
[root@localhost ~]# tar -cvf anaconda-ks.cfg.tar anaconda-ks.cfg
#把anacondhks.cfg打包为 anacondhks.cfg.tar文件
```

选项“-cvf”一般是习惯用法，记住打包时需要指定打包之后的文件名，而且要用“.tar”作为扩展名。打包目录也是如此：

```
[root@localhost ~]# ll -d test/
drwxr-xr-x 2 root root 4096 6月 17 21:09 test/
#test是我们之前的测试目录
[root@localhost ~]# tar -cvf test.tar test/
test/
test/test3
test/test2
test/test1
#把目录打包为test.tar文件
tar命令也可以打包多个文件或目录，只要用空格分开即可。例如：
[root@localhost ~]# tar -cvf ana.tar anaconda-ks.cfg /tmp/
#把anaconda-ks.cfg文件和/tmp目录打包成ana.tar文件包
```

【例 2】打包并压缩多个文件。

```
[root@hadoop101 opt]# tar -zcvf houma.tar.gz houge.txt bailongma.txt
houge.txt
bailongma.txt
[root@hadoop101 opt]# ls
houma.tar.gz houge.txt bailongma.txt
```

【例 3】打包压缩目录。

```
[root@hadoop101 ~]# tar -zcvf xiyou.tar.gz xiyou/
xiyou/
xiyou/mingjie/
xiyou/dssz/
xiyou/dssz/houge.txt
```

4.2 tar命令做解打包操作

当 tar 命令用于对 tar 包做解打包操作时，该命令的基本格式如下：

```
[root@localhost ~]#tar [选项] 压缩包
```

当用于解打包时，常用的选项与含义如表 2 所示。

选项	含义
-x	对 tar 包做解打包操作。
-f	指定要解压的 tar 包的包名。
-t	只查看 tar 包中有哪些文件或目录，不对 tar 包做解打包操作。
-C 目录	指定解打包位置。
-v	显示解打包的具体过程。

其实解打包和打包相比，只是把打包选项 "-cvf" 更换为 "-xvf"。我们来试试：

```
[root@localhost ~]# tar -xvf anaconda-ks.cfg.tar
#解打包到当前目录下
```

如果使用 "-xvf" 选项，则会把包中的文件解压到当前目录下。如果想要指定解压位置，则需要使用 "-C(大写)" 选项。例如：

```
[root@localhost ~]# tar -xvf test.tar -C /tmp
#把文件包test.tar解打包到/tmp/目录下
```

如果只想查看文件包中有哪些文件，则可以把解打包选项 "-x" 更换为测试选项 "-t"。例如：

```
[root@localhost ~]# tar -tvf test.tar
drwxr-xr-x root/root 0 2016-06-17 21:09 test/
-rw-r--r- root/root 0 2016-06-17 17:51 test/test3
-rw-r--r- root/root 0 2016-06-17 17:51 test/test2
-rw-r--r- root/root 0 2016-06-17 17:51 test/test1
#会用长格式显示test.tar文件包中文件的详细信息
```

