

Compressed text files

GETTING CLOSER

Gzip and bzip2 not only compress files, they also provide lean and powerful tools for viewing, searching, and comparing text files. **BY HEIKE JURZIK**

Compressed files save space not only in data transfers, but also on your local hard disk. The most frequently used compression tools are gzip and bzip2. The packages include numerous tools for ongoing use of packed data. Whether you just need a quick glance or want to read a file thoroughly, search a file for patterns, or compare files, Linux has the right tools out of the box.

The cat program outputs text files and more to standard output without the need to unpack compressed files. Zcat

relies on gzip and bzip2 relies on bzip2 to display packed text files on screen:

```
$ zcat /var/log/apache2/
access.log.10.gz
```

Zcat and bzip2 do not modify the original file. Although you might expect them to be able to do so, neither of the tools can be used to concatenate compressed files.

The command

```
$ zcat txt1.gz txt2.gz >
>> txt3.gz
```

does not return a compressed file named *txt3.gz*. Instead, as calling *file txt3.gz* reveals, it returns a normal unpacked text file. The same thing applies to bzip2, so if you want to glue compressed files together, you will need the standard cat program (Figure 1).

If you want to view a longer text file on screen, scroll through the file, or search for something inside the file, cat is not the right choice. Instead, you should use a pager, such

as more or less. The counterparts for packed text files are zmore/zless or bzip2more/bzip2less.

All the keyboard shortcuts that work with the two pagers are supported here, too. For example, you can scroll by pressing the space or arrow keys; pressing *H* displays the help text, and *Q* quits the program. The environment variables *\$MORE* and *\$LESS* are also interpreted by the tools and let you define standard behavior (i.e., the default command-line options) for the pagers.

The grep tool is useful for finding words and expressions in text files. If the files happen to be compressed gzip and bzip2 files, zgrep or bzgrep offers a practical solution. Both tools understand grep's command-line parameters, such as *-color* (highlight matches), *-i* (ignore case), *-r* (recursive search through directory trees), and so on.

Another practical consideration is that both tools evaluate the grep environment variables. For example, you can use *GREP_OPTIONS* to define the tool's default behavior or *GREP_COLOR* to set your own colors, and both zgrep and bzgrep follow suit.

Cmp and diff are used to compare files in the shell. Their counterparts for compressed files are the zcmp/bzcmp and zdiff/bzdiff tools (Figure 2).

All four programs understand the usual diff and cmp parameters and compare two files passed in to them. ■

```
Shell - Konsole
Session Edit View Bookmarks Settings Help
knoppix@Knoppix:~$ cat txt1.bz2 txt2.bz2 >> txt3.bz2
knoppix@Knoppix:~$ file txt3.bz2
txt3.bz2: bzip2 compressed data, block size = 900k
knoppix@Knoppix:~$
```

Figure 1: Follow the standard procedure to concatenate compressed files with the cat command.

```
Shell - Konsole <3>
Session Edit View Bookmarks Settings Help
knoppix@Knoppix:~$ bzdiff txt4.bz2 txt5.bz2
1,3c1
< Well maybe I test bzdiff now.
<
<
---
> Well maybe me test bzdiff now.
knoppix@Knoppix:~$
```

Figure 2: With zdiff or bzdiff, you can easily find differences between different versions of compressed files.