Kyiv College of Communications

Cycle Commission of Computer Engineering

**PERFORMANCE REPORT**

**LABORATORY WORK №5**

*in discipline: "Operating systems"*

*Topic: "Linux commands for archiving and compressing data"*

Performed by students

RPZ-93A group

Makarenko Danylo

Nemechkin Maxim

The teacher checked

Povkhlib V.S.

Kyiv 2021

**The goal of the work:**

1. Gaining practical skills in working with the Bash command shell.

2. Familiarity with the basic actions when working with help.

3. Familiarity with the basic actions when working with files and directories.

**Material support of classes**

1. Computer type IBM PC.

2. Windows family of operating systems (Windows 7).

3. Virtual machine - Virtual Box (Oracle).

4. GNU / Linux operating system - CentOS.

5. Cisco Network Academy website netacad.com and its online Linux courses

Tasks for preliminary preparation.

**The material was prepared by student Niemiechkin M.D.**

1.

|  |  |
| --- | --- |
| instance | екземпляр |
| distribution | розповсюдження |
| opportunities | можливості |
| compression techniques | техніки стиснення |
| reconstruct the original file | відновити вихідний файл |
| placeholder | заповнювач |
| implements | знаряддя |
| measurement of memory | вимірювання пам'яті |

**The material was prepared by student Makarenko D.A.**

2.1. What is the purpose of the commands tar, xz, zip, bzip, gzip? Make a brief description of each team and highlight their main parameters. How to install them.

Gzip is a file compression and recovery utility that uses the Deflate algorithm.

According to the tradition of UNIX programming, gzip performs only one function: compressing and unpacking one file, it can not pack multiple files into one archive. When compressed, the .gz suffix is ​​added to the original file extension. To package multiple files, they are usually first archived into one file with the tar utility, and then this file is compressed by gzip. Thus, compressed archives usually have a double extension .tar.gz.

In the simplest case, it is called in the following format:

[user] $ gzip file

Tar is the only archive of a collection of files and folders - but unlike zip, tar is not compressed.

Create a tar file containing all your images, preserving the folder structure, using the following command: tar -cvf photos ~ / photos

Bzip2 is a free, open source command-line utility (as well as an algorithm) for compressing data.

bzip2 performs only one function: compressing and unpacking a single file. The .bz2 extension is added to the default file name. To package multiple files, they are usually first archived into one file with the tar utility, and then compressed with bzip2. Such archives usually have the extension .tar.bz2.

bzip2 compresses most files more efficiently, but slower than more traditional gzip or zip. \

bzip2 can be used both in combination with tar and separately: bzip2 file for compression and bzip2 -d file.bz2 for unpacking

Xz is a lossless computer compression program that uses the LZMA2 algorithm as well as the appropriate file format.

xz compresses a single file on the input, rather than fitting a set of files into a single archive. Thus, it compresses a file, which can also be an archive, such as those created by Unix programs such as tar or cpio.

Zip

In Linux, it is easy to use the Zip (for creation) and Unzip (for extension) archives.

To check the installation, type "sudo apt-get install zip unzip" (without quotes)

Once installed, we can use zip to create archives (or modify existing ones), and unzip to expand them to the originals.

2.2. Here are three examples of how to archive and compress data with different commands.

bzip2 compresses most files more efficiently, but slower than more traditional gzip or zip

gzip To package multiple files, they are usually first archived in one file with the tar utility, and then this file is compressed by gzip

zip The easiest way to use zip is to tell it the name of the zip archive you want to create, and then explicitly name each file that should be included in it.

Test questions

**The material was prepared by student Nemechkin M.D.**

1. ***Give a comparative description of compression and archiving processes.***

Data compression is a procedure of recoding data in order to reduce its volume.

Archiving of information - compression of one or more files in order to reduce the amount of data when stored on media or when transmitting data over communication channels, including the Internet, and placing compressed files in one archive file.

When data compression methods are applied to finished files, the term "data compression" is often used instead of "data compression", the compressed version of the data is called an archive, and software that implements compression methods are called archivers.

Depending on the object in which the data to be compressed is located, there are:

1. Compression (archiving) of files: used to reduce the size of files in preparation for transmission through communication channels or for transportation on external low-capacity media;

2. Compression (archiving) of folders: used as a means of reducing the volume of folders before long-term storage, for example, when backing up;

3. Compression (compression) of disks: used to increase the efficiency of disk space by compressing data when writing them to media (usually by means of the operating system).

1. ***What other applications can be used to compress and archive Linux files and directories? Give examples and a brief description.***

Ark is a KDE archiving tool included in the Kdeutils package.

File Roller is a computer archiver for the GNOME desktop. File Roller is a graphical shell that provides a single graphical interface for console archiving programs. Distributed under the GNU General Public License.

ARJ is a file archiver. Developed by Robert K. Jung. (Origin of the name ARJ: Archiver Robert Jung). ARJ version 1.00 was released in February 1991 under a shareware license. ARJ compression is similar to PKZIP 1.02 There is also an open source version of ARJ, available under more than ten operating systems, including various versions of UNIX and Linux. There is also a version of Russian NLV, which allows you to protect archives with encryption algorithm GOST.

bzip2 is a free, open source command-line utility (as well as an algorithm) for compressing data. Developed and first published by Julian Seward in July 1996 (version 0.15). The stability and popularity of the compressor grew over the years, and version 1.0 was released in late 2000. According to UNIX programming traditions, bzip2 performs only one function: compressing and unpacking a single file.

gnochive - GUI for GNOME archivers.

Easibox - Console utility for creating various archives.

Lha - File Archiver - uses the LZW compression method (.lzh files).

kArchiver - Advanced archiver, understands tar, tgz, tar.gz, bz2, zip, rar and others.

mediArchiver - Media archiver for storing all the richness of media files.

Compare the compression algorithms used in the commands (programs) used in Linux.

1. ***Which of the algorithms can be considered the fastest and most efficient?***

Two types of methods are used - lossless and lossless compression algorithms. The first allows you to restore the file to its original state without losing a single bit of information in a compressed file. The second is a typical approach to executables, text and spreadsheets, where the loss of words or numbers will change the information.

D. Huffman's algorithm and Shannon-Fano's algorithm involve recoding data. Characters that are more common in a message are recoded by fewer characters and less often by more characters. For example, to encode in Windows-1251 the previous sentence, which contains 114 characters, including spaces, you must use 1 byte of data for each character, or 114 bytes. If you count how many times a character occurs in a sentence, it turns out that the total number of characters in the sentence - 28, the most common symbol "space" - 14 times, the symbol "i" - 10, "o" - 9, "c" - 7, "c" - 6, etc. At least once, there are symbols "shch", "b", "c", "-", ".". If you use the data compression algorithm for binary code (0 and 1) and mark the most common characters with one character (1 bit): the character "space" - 0, and the character "and" - 1, followed by the frequency of repetition - two characters. 2 bits): character "o" - 00, character "c" - 01, etc. For less common characters, four binary digits or 4 bits will be used. As a result, the total code length of the sentence encoded in this way will be 90 bits or 11 bytes and 2 bits. As you can see, the total amount of sentence code will decrease 10 times.

Of course, this is a fairly simplified coding scheme, do not forget that the file compressed in this way should contain a coding table and additional service information. Therefore, when encoding very small files, there may be no reduction, but, on the contrary, increase in file size. For files with a large amount of text or digital data, this algorithm provides a high degree of compression.

The Run-Length encoding (RLE) algorithm is more commonly used to encode graphics data. This method involves replacing a sequence of similar data with its abbreviated description. For example, the image of the Apple logo (Fig. 2.49) uses only two colors - black and white. If you mark black with the number "1" and white with "0", then the code of the first line of the image should consist of 48 digits "0" (48 bits).

But they can be replaced by a description, specifying the symbol code and the number of repetitions - 0 48. The description code will be allocated 9 bits: 1 bit to the digit "0" and four bits to two digits of the number of repetitions. The second row of pixels will be encoded in the same way, and the third: 0 29 1 4 0 15-23 bits. For color images with a large number of monochrome fragments, this algorithm will provide high compression efficiency.

1. ***Describe the compression and archiving software that can be used on your mobile phone.***

Nowadays, we come across archive files very often on all platforms: Windows, Mac or Linux

The most popular Linux archiving utility is tar. It is used almost everywhere, for archiving source code, packing packages. Many other programs are used for compression, depending on the compression algorithm, such as zip, bz, xz, lzma, etc.

1. ***Describe and compare software for compressing and (de) archiving data in the Windows family of operating systems.***

There are two types of archivers:

Archivers running in command line mode. The essence of working with these archivers is to enter commands that match the name of the executable file of the program. It is very important not to forget about setting the appropriate parameters.

Shell archivers are programs with a user-friendly interface that makes it easy for any user to perform archive operations.

Data compression is the process of transcoding data to reduce file size.

1. ***Explain how data compression and archiving can be used to back up data. In what other tasks of system administration it can be used.***

It is used for more efficient use of storage and data transmission devices. Compression is based on the elimination of excess contained in the original data. The simplest example of redundancies is the repetition of fragments in the text. Problems such as data duplication, data transfer and work with general documents are solved.

**Conclusion**

**The material was prepared by student Makarenko D.A.**

During the laboratory work, we gained practical skills in working with the Bash command shell. We also learned the basic commands for archiving and compressing data.