“Київський фаховий коледж зв’язку”

Циклова комісія комп’ютерної та програмної інженерії

**ЗВІТ ПО ВИКОНАННЮ**

**ЛАБОРАТОРНОЇ РОБОТИ №6**

з дисципліни: «Операційні системи»

**Тема: «Команди Linux для архівування та стиснення даних. Робота з текстом»**

Виконали студенти

групи КСМ-23А

Команда КГК:

Корольов Є.Ю.,

Горохов Д.С. та

Коваленко С.О.

Перевірила викладач

Сушанова В.С.

Київ 2024

**Мета роботи:**

1. Отримання практичних навиків роботи з командною оболонкою Bash.
2. Знайомство з базовими командами для архівування та стиснення даних.
3. Знайомство з базовими діями при роботі з текстом у терміналі.

**Матеріальне забезпечення занять**

1. ЕОМ типу IBM PC.

2.  ОС сімейства Windows та віртуальна машина Virtual Box (Oracle).

3. ОС GNU/Linux (будь-який дистрибутив).

4. Сайт мережевої академії Cisco netacad.com та його онлайн курси по Linux.

**Завдання для попередньої підготовки:**

***Готував матеріал студент Корольов Євгеній.***

1. \*Прочитайте короткі теоретичні відомості до лабораторної роботи та зробіть невеликий словник базових англійських термінів з питань призначення команд та їх параметрів.

|  |  |
| --- | --- |
| Термін англійською | Термін українською |
| **cd** | |  | | --- | | змінити директорію |  |  | | --- | |  | |
| |  | | --- | | **ls** |  |  | | --- | |  | | |  | | --- | | показати вміст директорії |  |  | | --- | |  | |
| **pwd** | |  | | --- | | показати поточну директорію |  |  | | --- | |  | |
| **cp** | |  | | --- | | копіювати файл/папку |  |  | | --- | |  | |
| **mv** | перемістити або перейменувати файл/папку |
| **rm** | видалити файл/папку |
| |  | | --- | | **mkdir** |  |  | | --- | |  | | створити нову директорію |
| **rmdir** | |  | | --- | | видалити порожню директорію |  |  | | --- | |  | |
| |  | | --- | | **touch** |  |  | | --- | |  | | |  | | --- | | створити новий файл |  |  | | --- | |  | |
| **echo** | |  | | --- | | вивести текст у термінал |  |  | | --- | |  | |
| **kill** | |  | | --- | | завершити процес |  |  | | --- | |  | |
| |  | | --- | | **top** |  |  | | --- | |  | | |  | | --- | | показати список процесів у реальному часі |  |  | | --- | |  | |
| **sudo** | |  | | --- | | виконати команду з правами адміністратора |  |  | | --- | |  | |

1. На базі розглянутого матеріалу дайте відповіді на наступні питання:
   1. \*Яке призначення команд *tar*, *xz*, *zip*, *bzip*, *gzip*? Зробіть короткий опис кожної команди та виділіть їх основні параметри. Яким чином їх можна встановити.

* **Tar** is one of the most widely used Linux compression commands.

Advantages of tar:

* Tar compresses files quite effectively, the compression ratio is 50%;
* Significantly reduces the size of packed files and folders;
* Tar does not change the properties of files and directories. Permissions and other features remain unchanged;
* Tar is available in most versions of Linux, including older ones, as well as in Android firmware;
* Quickly compresses and decompresses content;
* Easy to use.

**XZ** is a compressed file format that uses the LZMA2 compression algorithm. It was developed as a replacement for the popular gzip and bzip2 formats and offers a number of advantages over these older standards. XZ files are well supported on many platforms and can be decompressed quickly and easily. Although not as common as ZIP or RAR files, XZ archives can be used to store large amounts of data without sacrificing compression quality. If you need to compress or decompress large files, the XZ file extension is worth considering.

To unzip XZ files, first install the xz-utils package:

$ sudo apt-get install xz-utils

**ZIP** is a compression and file packaging utility for Unix. Each file is stored in a single .zip {.zip-filename} file with the extension .zip.

* Zip is used to compress files to reduce file size and is also used as a file package utility. Zip is available in many operating systems like Unix, Linux, windows, etc.
* If you have limited bandwidth between two servers and want to transfer the files faster, then zip the files and transfer them.
* The zip program puts one or more compressed files into a single zip archive, along with information about the files (name, path, date, time of last modification, protection, and check information to verify file integrity). An entire directory structure can be packed into a zip archive with a single command.
* Compression ratios of 2:1 to 3:1 is common for text files. zip has one compression method (deflation) and can also store files without compression. zip automatically chooses the better of the two for each file to be compressed. The program is useful for packaging a set of files for distribution, archiving files, and for saving disk space by temporarily compressing unused files or directories.

**bzip2** command in Linux is used to compress and decompress the files i.e. it helps in binding the files into a single file which takes less storage space than the original file used to take. It has a slower decompression time and higher memory use. It uses Burrows-Wheeler block sorting text compression algorithm, and Huffman Coding. Each file is replaced by a compressed version of itself, with the name original name of the file followed by the extension bz2.

Syntax:

bzip2 [OPTIONS] filenames ...

**Gzip** is one of the most popular compression algorithms that allows you to reduce the size of a file and keep the original file mode, ownership, and timestamp. Gzip is the algorithm that is used by web servers to compress pages and assets before sending it to the client browser.

Gzip also refers to the .gz file format and the gzip utility , which you can use to compress and decompress files.

This article will teach you how to use the gzip command.

Basics of gzip:

Gzip, short for GNU Zip, is a command-line compression tool commonly found on Linux systems. It utilizes the DEFLATE compression algorithm to reduce the size of files, making them more manageable for storage and transmission.

* 1. \*\*Наведіть три приклади реалізації архівування та стискання даних різними командами.

1. tar -czf archive.tar.gz /path/to/directory — creates a compressed archive.tar.gz with the contents of the directory using tar and gzip;

2. zip -r archive.zip /path/to/directory — creates a compressed archive.zip, adding files recursively inside the directory;

3. tar -cf - /path/to/directory | gzip > archive.tar.gz is an alternative way to create a tar.gz archive through the standard stream, compression is done in gzip.

* 1. \*Яке призначення команд cat, less, more, head and tail? Зробіть короткий опис кожної команди та виділіть їх основні параметри. Яким чином їх можна встановити

**1. cat**

Purpose: The cat command is used to output the contents of files to the standard output (screen) or to combine several files.

Main parameters:

-n: number lines.

-b: number only non-empty lines.

-E: mark end of line with $ character.

Installation: Usually, cat is already installed in most Linux distributions.

**2. less**

Purpose: The less command allows you to view the contents of files by pages, which makes it possible to scroll the text up and down.

Main parameters:

-N: show line numbers.

-S: Trim long lines so they don't go off the screen.

-F: exit if content fits on one screen.

Installation: Less is usually already installed in most distributions. If not, it can be installed via the package manager, for example:

bash

Copy the code

sudo apt install less # for Ubuntu

sudo yum install less # for CentOS/RHEL

**3. more**

Purpose: The more command is also used to view the contents of files, but with fewer options than less. It shows the text in parts by scrolling down.

Main parameters:

-n: specify the number of lines to review before moving to the next part.

-c: clear the screen before displaying a new page.

Installation: Usually more is already installed. If not, it can also be installed via the package manager:

bash

Copy the code

sudo apt install more # for Debian/Ubuntu

**4. head**

Purpose: The head command is used to output the first few lines of a file.

Main parameters:

-n <number>: specify the number of lines to output (default is 10).

-c <number>: print the specified number of bytes.

Installation: head is usually included in the base utilities package in Linux.

**5. tail**

Purpose: The tail command outputs the last few lines of a file.

Main parameters:

-n <number>: specify the number of lines to output (default is 10).

-f: watch for new entries in the file in real time.

Installation: tail is also part of the basic set of utilities in Linux..

* 1. \*\*Поясніть принципи роботи командної оболонки з каналами, потоками та фільтрами

**1. Streams**

In the shell context, streams are abstractions used for input and output. The main types of flows:

• Standard input (stdin): This is the stream from which the program receives input (usually the keyboard).

• Standard output (stdout): This is the stream where the program outputs its results (usually the screen).

• Standard error output (stderr): This is the stream where the program outputs error messages (also usually the screen).

**2. Channels**

Channels allow you to connect the output of one process with the input of another. This is done using the symbol | (pipe).

This allows you to combine commands, process data without creating intermediate files, and reduces the need for manual data entry.

**3. Filters**

Filters are programs that accept input, process it, and output results. They usually read data from standard input and write the results to standard output. Some popular filters include:

• grep: to find text strings that match a given pattern.

• sort: to sort strings.

• awk: for processing and formatting text data.

• sed: for text editing.

* 1. \*Яке призначення команди grep?

A command used to filter lines in a file or the output of another command that match a given pattern.

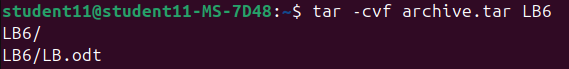
**Хід роботи:**

***Готував матеріал студент Корольов Є та Горохов Д(401 ауд.)***

* 1. Початкова робота в CLI-режимі в Linux ОС сімейства Linux:
  2. Запустіть операційну систему Linux Ubuntu. Виконайте вхід в систему та запустіть термінал ***(якщо виконуєте ЛР у 401 ауд.)***.
  3. Запустіть віртуальну машину Ubuntu\_PC ***(якщо виконуєте завдання ЛР через академію netacad)***
  4. Запустіть свою операційну систему сімейства Linux ***(якщо працюєте на власному ПК та її встановили)*** та запустіть термінал.
  5. Опрацюйте всі приклади команд, що представлені у лабораторних роботах курсу ***NDG Linux Essentials - Lab 9: Archiving and Compression*** та ***Lab 10: Working With Text.*** Створіть таблицю для опису цих команд

|  |  |
| --- | --- |
| Назва команди | Її призначення та функціональність |
| mkdir mybackups | Creating a new mybackups directory in the user's home directory |
| tar -cvf mybackups/udev.tar /etc/udev | The tar command is used to combine multiple files into a single file. In this case, the contents of the /etc/udev directory will be saved in the udev.tar archive in the mybackups directory. The -c option tells the tar command to create a tar file. The -v option stands for "verbose", which tells the tar command to demonstrate what it is doing. The -f option is used to specify the name of the tar file. |
| bzip2 | Command used to compress and decompress files using the Burrows-Wheeler compression algorithm. |
| gzip | Command used to compress and decompress files using the Lempel-Ziv compression algorithm. |
| tar | Command used to create, extract, and view archive files. The traditional UNIX archiving utility. |
| unzip | Command used to list and extract files from ZIP archives. |
| xz | Command used to compress and decompress files using the Lempel-Ziv-Markov compression algorithm. |
| zip | Command used to archive and compress files. The primary archiving utility in Microsoft. |
| tar -tjf folders.tbz | -t (List the files in an archive),-j (Decompress with an bzip2 command), -f ARCHIVE (Operate on the given archive). |
| tar -x [-f ARCHIVE] [OPTIONS] | -v(Verbosely list the files processed). |
| zip [OPTIONS] [zipfile [file…]] | The first argument zipfile is the name of the archive to be created, after that, a list of files to be added. |
| cat | The cat command, short for *concatenate*, is a simple but useful command whose functions include creating and displaying text files, as well as combining copies of text files. One of the most popular uses of cat is to display the content of text files. |
| Cut | Command used to extract columns of text from a file or standard input. |
| grep | Command used to filter lines in a file or the output of another command that matches a specified pattern. |
| head | Command used to display only the first few lines of a file. |
| less | Pager command used to view file contents one page of data at a time. |
| sort | Command used to rearrange lines of files or input. |
| tail | Command used to display only the last few lines of a files. |
| wc | Command that displays statistics of files such as the number of lines, words, and bytes. |

* 1. Ознайомтесь з командою tar та за її допомогою виконати у терміналі наступні дії:
* створити файл з розширенням .tar;



* створити файл з розширенням .tar, що складається з декількох файлів і каталогів одночасно;



* перегляду вмісту файлу;



* витягти вміст файлу tar;



* створити архівний файл tar, стиснений за допомогою bzip;



* витягти вміст файлу tar bzip; 
* створити архівний tar файл, стисненого за допомогою gzip;



* витягти вміст файлу tar gzip.



* 1. \*Як буде відбуватись перенаправлення потоків виведення в bash для наступних дій з командами (позначено як cmd) та файлами (позначено як file):

|  |  |
| --- | --- |
| **Команда** | **Що виконує команда?** |
| cmd 1> file | Перенаправляє стандартний потік в файл file |
| cmd > file | Відповідний функціонал до першого |
| cmd 2> file | Відповідний функціонал до першого |
| cmd >> file | Додає вміст до файлу |
| cmd &> file | Всі дані та помилки будуть записані в файл |
| cmd > file 2>&1 | Спочатку перенаправляє стандартний потік, потім помилки записані в файл |
| cmd >> file 2>&1 | Додає вміст до файлу, потім помилки записує в файл |
| cmd 2>&1 > /dev/null | Усі помилки будуть відображені в терміналі |
| cmd 2> /dev/null | Стандартний вивід залишається в терміналі, але всі помилки ігноруються |
| cmd1 | cmd2 | Вивід першої команди використовується як вхід для другої команди |
| cmd1 2>&1 | cmd2 | Всі дані, включаючи помилки, будуть передані у cmd2 |

* 1. \*\*Розгляньте наведені нижче приклади та поясніть, що виконують дані команди та який тип перенаправлення потоків вони використовують:

|  |  |  |
| --- | --- | --- |
| **Команда**  **(контейнер команд)** | **Що виконує команда?** | **Який потік перенаправлення?** |
| $echo "It is a new story." > story | Створює або перезаписує файл story текстом "It is a new story." | Перенаправляє стандартний вивід команди echo у файл. Якщо файл уже існує - його вміст буде перезаписаний |
| $ date > date.txt | Виводить поточну дату і час та зберігає їх у файл date.txt | Перенаправляє стандартний вивід команди date у файл date.txt. Якщо файл уже існує - його вміст буде перезаписаний |
| $ cat file1 file2 file3 > bigfile | З’єднує вміст файлів file1 file2 file3, а потім записує цей вміст у файл bigfile | Перенаправляє стандартний вивід команди cat у файл bigfile. |
| $ls -l >> directory | Виводить довгий список файлів і каталогів в поточній директорії і додає цей вивід до файлу directory | Використовується для додавання виводу до кінця файлу directory без перезапису існуючого вмісту |
| $ sort < file1\_unsorted > file2\_sorted | Читає вміст файлу file1\_unsorted, сортує його і записує результат у файл file2\_sorted | < перенаправляє вхід з файлу  file1\_unsorted  > перенаправляє вивід команди sort у файл file1\_unsorted |
| $ find -name '\*.txt' > file.txt 2> /dev/null | Шукає всі файли з розширенням .txt у поточному каталозі та записує результат у файл file.txt. Помилки ігноруються і перенаправляються у /dev/null | > перенаправляє стандартний вивід у файл file.txt  2> /dev/null/ перенаправляє потік помилок у /dev/null, видаляючи його. |
| $ cat file1\_unsorted | sort > file2\_sorted | Читає вміст файлу file1\_unsorted, сортує його за допомогою sort, а потім записує результат у файл file2\_sorted. | | передає стандартний вивід команди cat у вхід команди sort  > перенаправляє вивід команди sort у файл file2\_sorted |
| $ cat myfile | grep student | wc -l | Читає вмісту файлу myfile, знаходить всі рядки, що місять слово student, та піраховує кількість таких рядків. | | передає вивід команди cat myfile у вхід команди grep student  Друга | передає результат grep у вхід команди wc -l, яка рахує кількість рядків. |

**Контрольні запитання:**

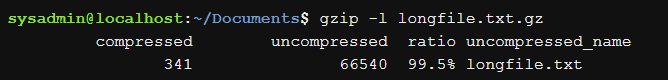
***Виконала матеріал студентка Коваленко С. О.***

1. Надайте порівняльну характеристику процесам стискання та архівування.

Compression - the use of algorithms that eliminate redundant information, thus reducing the file size.

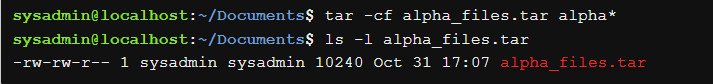
Lossy: some information is removed to significantly reduce the size.

Lossless: no information is lost, i.e. compression allows you to restore the original file exactly.



The idea of “archiving” data usually means backing it up and storing it in a safe place, often in a compressed format. On a Linux server, a program called tar was created that allows you to access and manipulate many files on a file system, while maintaining permissions and metadata as if they were a single file. You can then extract the file or the entire file system from the archive.

Essentially, a tar file is a file format that creates a convenient way to distribute, store, backup, and manipulate groups of related files.



1. Які програми, окрім наведених в роботі, можуть використовуватись для стискання та архівування файлів та каталогів в ОС Linux? Наведіть приклади та їх короткий опис.

zstd is a lossless compression algorithm that has been in development since 2015 with the support of Facebook. zstd was designed to provide compression and decompression speeds while maintaining a compression ratio comparable to the Deflate algorithm (which was developed in 1991 and used in the original ZIP and gzip programs).



LZ4 is a lossless data compression algorithm focused on high compression and decompression speeds. It belongs to the LZ77 family of compression methods that work with byte streams. It is characterized by a compact code for decompression.

The LZ4 algorithm has a slightly lower compression ratio than the previous LZO compression method. LZO, in turn, compresses at a lower rate than the classic gzip and DEFLATE . However, LZ4 is close to LZO in terms of compression speed and several times faster than gzip, and LZ4's decompression speed is much faster than LZO.

arj is an archiving utility that has been popular since the 90s. It supports the creation of multi-volume archives.

1. \*Порівняйте алгоритми стискання, що використовуються в командах (програмах), використовуваних в Linux. Які з алгоритмів можна вважати найшвидшим та найефективнішим?

gzip has a compression rate of medium. The unpacking speed is high. The compression rate is medium. The resource usage is low. Very popular and compatible with many programs. A good choice for fast file compression, especially for large amounts of data that are often decompressed. Not as efficient as newer algorithms (e.g., zstd or xz) in terms of maximum compression.

bzip2 has a lowcompression rate (much slower than gzip). The decompression speed is average. The compression ratio is high (better than gzip). Resource usage is high during compression. The high compression ratio makes it useful for archiving large files when the priority is to keep the size to a minimum. Slow compression, which makes it less suitable for cases where speed is important.

xz has a lowcompression speed (slower than bzip2). The decompression speed is medium. Compression ratio is very high (better than bzip2 and gzip). Resource usage: high, especially memory during compression. Maximizes file size, making it suitable for backup and archiving. Very slow compression process, which can be a problem for large amounts of data. Requires a significant amount of RAM.

zstd has a highcompression speed (faster than gzip and xz). The decompression rate is very high (faster than gzip). Compression ratio: medium to very high (depending on the settings). Resource usage is average, but flexible settings allow you to adjust resource usage. A compromise between compression speed and quality. Supports different compression levels, allowing you to customize the settings as needed. It is the best choice for big data where speed is of the essence. At the maximum compression levels, it may still be slower than gzip, but the performance is better.

lz4 has a compression rate of very high (the fastest among the listed). The decompression speed is very fast. The compression ratio is low (worse than gzip and zstd). Resource usage is low. Very fast and lightweight for real-time or log processing applications. Ideal for cases where speed is more important than compression ratio. Less efficient than other algorithms in reducing file size, making it less suitable for archiving where maximum compression is important.

The fastest can be considered lz4 - it provides very fast compression and unpacking, but with a lower compression ratio.

The most efficient in terms of the balance between speed and compression ratio is zstd. It offers flexible settings that allow you to choose both fast compression and maximum file size reduction.

1. \*Опишіть програмні засоби для стискання та архівування, що можуть бути використані у вашому мобільному телефоні.

I have an iPhone cell phone.

If you need basic compression and decompression of .zip files, the built-in Files app is the most convenient solution, as it's already installed on your iPhone.

If you need to work with a variety of archive formats and advanced features such as encryption, iZip and WinZip are good options.

If you need to work with .rar archives or other specific formats, RAR for iOS is a good choice.

Documents by Readdle is suitable for those who need a comprehensive file manager with additional features for working with archives.

1. \*Опишіть та порівняйте програмні засоби для стискання та (де)архівування даних у ОС сімейства Windows.

Windows has built-in support for creating and extracting .zip archives through Windows Explorer.

This allows you to compress files and folders to .zip without any additional software.

WinRAR isone of the most popular programs for compressing and decompressing files on Windows. It supports a wide range of file formats: .rar, .zip, .7z, .tar, .gz and others. Although WinRAR is conditionally free (30-day trial period), users can continue to use it after the trial period expires.

7-Zip is afree and open-source archiver that supports many formats, including .7z, .zip, .rar, .tar, .gzip, .bzip2, and others. The .7z format provides a high degree of compression thanks to the use of the LZMA and LZMA2 algorithm.

PeaZipis a free and open-source archiver that supports over 150 archive formats, including .7z, .rar, .zip, .tar, .gzip, .bzip2 and others. It is especially suitable for users who want to have more control over the compression and archiving process.

For basic needs, if you only need to compress and decompress .zip files without installing additional programs, the built-in Windows function will work best.

For more complex archives and high compression, 7-Zip is the best free choice because it supports many formats and provides a high compression ratio in .7z.

For handling .rar files and ease of use, WinRAR remains the standard, especially if you deal with .rar archives frequently.

1. \*\*Поясніть яким чином стиснення та архівування даних може бути використано для резервування даних. В яких ще задачах системного адміністрування воно може бути використано.

Data compression and archiving are important tools in system administration and can be used for several purposes, including backup, storage, and transfer. When backing up data, system administrators often compress this data to reduce the size of the files. This saves disk space and reduces storage requirements. Archiving allows you to combine a large number of files into a single file. This simplifies backup management by allowing you to store structured data in a single container. Compression is useful when transferring large amounts of data over a network (for example, between servers or to transfer files to clients). Often, servers accumulate data that is no longer used regularly, but needs to be stored (for example, old event logs, database archives). The administrator can compress such data to reduce its size and store it in a compressed form on disk. Many compression and archiving tools allow you to add a password or encryption to the archive (for example, 7z, zip, rar). This allows you to store confidential backups or transfer them over a network with an additional layer of security.

1. \*\*Яке призначення директорії файлу /dev/null?

/dev/null is a virtual device file. Programs and utilities can interact with it as if it were a real file: request data, and send any information to it. But everything you write to /dev/null is sent to the void and forgotten (in other words, a black hole).

If you write any information to /dev/null, it simply disappears, leaving no trace.



Висновки:

In the course of the work, practical skills were acquired in working with the Bash shell, which is an important tool for managing Linux-based operating systems. Familiarization with the basic commands for archiving and compressing data allowed us to learn how to effectively create archives, compress files and unpack them for further use. This helps optimize disk space usage and facilitates file transfer. This significantly increases productivity when working with large amounts of data and automating processes. The knowledge gained allows you to use the command line capabilities to perform everyday tasks in Linux systems, which is an important step towards an in-depth understanding of administration and process automation in the environment of Unix-like systems.