### 3. Поиск по тексту. Работа с текстовыми файлами.

В своей основе Linux система с интерфейсом командной строки. Мы планируем применять инструменты командной строки в двух целях — манипуляции с файлами и каталогами и разбор файлов журналов. А иногда для простейших диагностических процедур.

Сегодня будем учится работать с текстовыми файлами.

Теория

1.

**В Linux многие, но не все сущности представлены в виде файлов.** Так например файлы в каталоге /dev - представляют различные физические или логические устройства. Например /dev/sda1 - первый раздел первого жёсткого диска.

Для нас практический интерес представляют **текстовые файлы**. Причина этого в том что **файлы журналов (логи или logs)** - почти всегда именно текстовые.

В Windows или DOS текстовый файл обычно отличается “расширением” (суффиксом) названия - .txt.

В Linux это не так. В Linux расширения не играют такой роли, хотя название файлов журналов обычно имеет окончание .log.

**Определение:**

**Текстовый файл содержит последовательность** [**символов**](https://ru.wikipedia.org/wiki/%D0%A1%D0%B8%D0%BC%D0%B2%D0%BE%D0%BB) **(в основном** [**печатных знаков**](https://ru.wikipedia.org/wiki/%D0%9F%D0%B5%D1%87%D0%B0%D1%82%D0%BD%D1%8B%D0%B5_%D1%81%D0%B8%D0%BC%D0%B2%D0%BE%D0%BB%D1%8B)**, принадлежащих тому или иному** [**набору символов**](https://ru.wikipedia.org/wiki/%D0%9D%D0%B0%D0%B1%D0%BE%D1%80_%D1%81%D0%B8%D0%BC%D0%B2%D0%BE%D0%BB%D0%BE%D0%B2)**). Эти символы обычно сгруппированы в строки (**[**англ.**](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) ***lines, rows*). В современных системах строки разделяются** [**разделителями строк**](https://ru.wikipedia.org/wiki/%D0%9F%D0%B5%D1%80%D0%B5%D0%B2%D0%BE%D0%B4_%D1%81%D1%82%D1%80%D0%BE%D0%BA%D0%B8)**. Различные операционные системы придерживаются своего представления перевода строки и конца файла. В** [**UNIX**](https://ru.wikipedia.org/wiki/UNIX) **перевод строки состоит из одного символа LF (код 0xA), в** [**Mac OS**](https://ru.wikipedia.org/wiki/Mac_OS) **(но не** [**OS X**](https://ru.wikipedia.org/wiki/OS_X)**) — из символа CR (код 0xD), а в** [**DOS**](https://ru.wikipedia.org/wiki/DOS) **и** [**Windows**](https://ru.wikipedia.org/wiki/Windows) **перевод строки кодируется последовательностью двух символов: CR и LF.**

<https://ru.wikipedia.org/wiki/%D0%A2%D0%B5%D0%BA%D1%81%D1%82%D0%BE%D0%B2%D1%8B%D0%B9_%D1%84%D0%B0%D0%B9%D0%BB>

Разбирая файлы журналов следует иметь в виду:

* Файлы журналов бывают огромны. Их размер может превышать размер оперативной памяти. Поэтому открывать файлы журналов в текстовом редакторе следует осторожно, убедившись что файл не превышает по размерам ОЗУ.
* Файлы журналов могут постоянно обновляться. В результате открытая в текстовом редакторе копия может не содержать актуальных записей.
* Файлы журналов могут быть сжаты архиватором. Например gzip-ом.

2.

Второе понятие которое нам потребуется - стандартный ввод-вывод и их перенаправление.

Когда вы запускаете в командной строке Linux (в Windows и DOS ситуация аналогичная) ту или иную утилиту (команду), чаще всего она выводит на экран текстовые сообщения. Например краткую информацию о себе, справку или результат своего выполнения.

Существует возможность **перенаправить вывод команды в текстовый файл**.

Например:

ls -lt / > /tmp/ls.txt

Создает **текстовый файл** /tmp/ls.txt с результатом выполнения команды ls -lt /

Этот текстовый файл можно просмотреть в Midnight Commander или иным образом.

**Символ перенаправления стандартного вывода команды: >**

Справедливо и обратное - если команда ожидает **ввода символов**, то возможно направить на её **стандартный ввод** содержимое текстового файла.

Например:

grep root < /etc/passwd

*С утилитой grep мы подробно познакомимся на следующем занятии.*

Найдёт в файле etc/passwd все строки содержащие слово “root”.

С утилитой grep мы сегодня ещё познакомимся.

**Символ перенаправления стандартного ввода команды: <**

И наконец третий вариант - **можно направить вывод одной команды на ввод другой команды - создать конвейер.**

Например:

ps -Al | grep tty

Выводит все процессы в описание которых есть строка tty.

**Символ перенаправления стандартного вывода на стандартный ввод другой команды: |**

Практика

1. Утилита less

Утилита less применяется для двух целей

* Просмотр текстовых файлов
* Постраничной прокрутки вывода других команд

Пример:

less /etc/passwd

Позволяет просмотреть файл /etc/passwd.

Пример:

ps -Al |less

Позволяет просмотреть информацию обо всех процессах в системе.

Выход из утилиты less - q (маленькая в английской раскладке).

2. Утилита tail

Утилита tail позволяет просматривать последние n строк файла.

Пример:

tail -n 2 /etc/group

Выводит последние 2 строки файла ./etc/group

3. Утилита head

Утилита head позволяет просматривать первые n строк файла.

Пример:

head -n 2 /etc/group

4. Утилита sort

Сортирует строки в файле по алфавиту в прямом или обратном порядке и не только.

Пример:

sort /etc/passwd

Выводит содержимое файла /etc/passwd отсортированное по алфавиту.

5. Утилита history

Что характерно выводит историю ранее введенных команд.

Пример:

History

Выводит все ранее введенные команды.

Также просмотреть ранее введенные команды можно при помощи клавиш - стрелка вверх и стрелка вниз.

6. Текстовый редактор vi

Взято: <http://lib.ru/unixhelp/vi.txt>

Очень кратко о редакторе "VI":

По не знаю какой уж там исторической традиции "честные"  
(канонические, особенно коммерческие) системы Unix не имеют в  
своем базовом комплекте ни одного нормального (еще раз  
подчеркну это слово - НОРМАЛЬНОГО) текстового редактора.  
Поэтому системный администратор должен уметь пользоваться  
тем, что ему дают. А дают ему несравненный "Визуальный" экранный  
редактор файлов - Редактор VI.  
  
Редактор VI имеет три режима:

Командный - в этом режиме можно перемещаться по файлу и  
выполнять корректирующие команды над текстом. Команды вызываются  
ОБЫЧНЫМИ ЛАТИНСКИМИ БУКВАМИ.  
  
Ввода текста - в этом режиме обычные латинские буквы будут  
вставляться в текст.  
  
Режим строчного редактора ED используется для управления  
файлами (типа сохранить файл, зачитать файл и т.д.)  
  
 Прочитайте предыдущий абзац еще раз. Вы еще не испугались?  
Хорошо. Поехали. Итак:

*vi имя\_файла*

VI в КОМАНДНОМ РЕЖИМЕ.  
  
ЧТОБЫ ВЫЙТИ ИЗ ФАЙЛА БЕЗ СОХРАНЕНИЯ, нажмите:  
 ESC : q ! Enter  
  
чтобы выйти из файла, сохранив изменения, нажмите:  
 ESC : w ! Enter  
 ESC : q Enter  
выйти из файла с сохранением, одной командой:  
 ESC : wq Enter  
  
для перехода В РЕЖИМ ВВОДА нужно нажать команды типа:

"i" вставлять здесь  
 "A" вставлять с конца строки  
 "cw" заменять текущее слово  
  
 ESC для ВОЗВРАТА В КОМАНДНЫЙ РЕЖИМ  
 CTRL-[ для возврата в командный режим  
  
 для перехода В РЕЖИМ УПРАВЛЕНИЯ ФАЙЛАМИ нужно нажать  
 ":" (перейти в режим редактора ED)

Перемещение по файлу:

h,j,k,l влево, вниз, вверх, вправо  
 Ctrl-F На страницу вниз  
 Ctrl-B На страницу вверх  
А если вам очень повезет, то можно будет двигаться стрелочками.  
Чтобы перейти в режим везения, нужно описать для операционной  
системы ваш терминал. Этим займемся в дальнейшем (см.  
["Описание терминалов, terminfo, termcap").](http://lib.ru/unixhelp/terminfo.txt)  
  
Подгоните курсор к нужному месту и нажмите  
 i перевод в режим ввода  
вводите требуемый текст  
 ESC прекратить ввод, перейти в командный режим  
  
Подгоните курсор к ненужному месту и нажмите  
 x удалить символ  
 dd удалить строчку  
  
Ещё несколько полезных комманд:

o вставлять с новой строки (под текущей строкой)  
 a в режим ввода ЗА курсором  
 5yy запомнить 5 строчек  
Подгоните курсор к нужному месту  
 p вставить запомненные строки под курсором  
 P вставить запомненные строки НАД курсором  
  
 J Склеить две строки  
 /Шаблон поиска Enter - поиск  
 n Повторить поиск  
  
На этом ознакомление с редактором VI можно считать законченным.

Задачи:

1. Прочитать man по утилите less
2. Прочитать man по утилите head
3. Прочитать man по утилите tail
4. Прочитать man по утилите sort
5. Прочитать man по редактору vi (основную часть)
6. Вывести все строки файла /etc/passwd содержащие слово bash результат направить в файл с именем bash.txt в своём домашнем каталоге
7. Вывести средствами tail весь файл /etc/fstab результат направить в файл с именем fstab.txt в своём домашнем каталоге
8. Вывести средствами head весь файл /etc/group результат направить в файл с именем head.txt в своём домашнем каталоге
9. Создать в домашнем каталоге при помощи редактора vi файл lear.txt и вставить в него фрагмент текста Короля Лир <https://en.wikisource.org/wiki/The_Tragedy_of_King_Lear>
10. Вставить в начало файла в задании 9 файла любую строку из man ls
11. Найти при помощи редактора vi в созданном файле строку со словами - “...This hideous rashness..” Сохранить найденную строку в отдельном текстовом файле в домашнем каталоге