

Отчет по лабораторной работе № 20 по курсу “Фундаментальная информатика”

Студент группы М80-103Б-21 Кудрявцев Андрей Георгиевич, № по списку 14

andr.kudriavtsev@gmail.com,
telegram: @AndrewBruhnaulsky

Работа выполнена: «11» марта 2022г.

Преподаватель: каф. 806 Севастьянов Виктор Сергеевич

Отчет сдан « » _____ 20__ г., итоговая оценка _____

Подпись преподавателя _____

Тема: Стандартные утилиты UNIX для обработки файлов

1. **Тема:** Стандартные утилиты UNIX для обработки файлов
2. **Цель работы:** Освоить основные утилиты UNIX и написать простой скрипт с их использованием.
3. **Оборудование:**
Процессор *Intel Core i5 1035G*, 4 ядра, 3.6GHz с ОП 8 Гб, НМД 1024 Гб. Монитор 1920x1080
4. **Программное обеспечение:**
 - Гипервизор – Virtualbox 6.1.26, ОП – 4 Гб
 - Операционная система семейства: *linux*, наименование: *ubuntu*, версия 21.04 Hisute Hippo
 - Интерпретатор команд: *bash* версия 5.1.4(1)
 - Система программирования – отсутствует
5. **Идея, метод, алгоритм:**
Опробовать все утилиты из следующего списка и запротоколировать их работу:

1	cmp	2	comm	3	wc	4	dd
5	diff	6	grep	7	join	8	sort
9	tail	10	tee	11	tr	12	uniq
13	od	14	sum	15	cut	16*	nroff
17*	vi/vim	18*	mc	19	tar	20	gzip
21	ed	22*	awk	23	sed	24	bzip2
25	head	26	iconv	27	patch	28	md5
29	du	30	file	31	touch	32	find
33	xargs	34	df	35	paste	36	cpp
37	indent	38	split	39	mktemp		

Освоив 20 и более утилит, напишем простой скрипт, используя утилиты *awk* или *sed*. Задача для скрипта возьмём из личного опыта.

Перед тем как сохранить изменения в *git* и опубликовать их на *GitHub* нужно убрать в переданном файле значения полей с конфиденциальными значениями.

6. Сценарий выполнения работы
Скрипт выполняет следующее:
1. Проверка того, что были заданы входной и выходной файл.
 2. Проверка того, что входной файл не пуст.
 3. Запись уникальных 5ых колонок (подстрок) входного файла в выходной файл.

Опробуем работу скрипта на /var/log/syslog:

```
andreii@UBU1A:~/labbl2/labab20$ ./script.sh /var/log/syslog out.txt
```

Результат (часть out.txt):

```
6 anacron[10255]:  
6 anacron[7929]:  
6 anacron[9182]:  
6 pulseaudio[1613]:  
6 systemd-timesyncd[12322]:  
8 mtp-probe:  
13 gnome-shell[1936]:  
15 PackageKit:  
18 update-notifier[5665]:  
19 avahi-daemon[544]:  
19 whoopsie[713]:  
33 packagekitd[2193]:  
64 snap-store[2459]:  
111 NetworkManager[549]:  
158 dbus-daemon[548]:  
170 snapd[7248]:  
172 snapd[10705]:  
172 snapd[7585]:  
174 snapd[10001]:  
181 dbus-daemon[1624]:  
182 snapd[6873]:  
190 snapd[7405]:  
200 snapd[9498]:  
208 kernel:  
212 snapd[9300]:  
214 snapd[9829]:  
227 snapd[579]:  
282 rtkit-daemon[1642]:  
297 systemd[1585]:  
345 systemd-resolved[6872]:  
618 systemd[1]:  
683 systemd-resolved[510]:  
959 gnome-shell[1797]:  
2616 systemd-resolved[7573]:  
"out.txt" 167L, 4214B
```

CMP – побайтовое сравнение файлов

```
andrei@UBU1A:~/labbl2/laba20$ cmp a.txt b.txt
a.txt b.txt differ: byte 8, line 3
```

```
q
w
e
r
t
yh
~
~
~
~
~
~
~
"a.txt" 6L, 15B
```

```
q
w
ea
rb
t
yh
~
~
~
~
~
~
~
"b.txt" 6L, 17B
```

COMM – построчное сравнение файлов (только 1/ только 2/общ)

```
andrei@UBU1A:~/labbl2/laba20$ comm a.txt b.txt
          q
          w
e
          ea
r
          rb
          t
          yh
andrei@UBU1A:~/labbl2/laba20$
```

WC – вывод кол-ва строк, слов, байтов файла

```
andrei@UBU1A:~/labbl2/laba20$ wc -c -l -w wc.txt
10 14 88 wc.txt
andrei@UBU1A:~/labbl2/laba20$
```

```
ewfedniojig weofivvevw
wefeswef
wef sgdfs f sdff
f ff ff f
uuu
qrwefgfdgws
```

```
gef
g
~
```

DD – побайтное копирование потока данных

```
andrei@UBU1A:~$ ls -lR | dd of=ls.txt
568+1 records in
568+1 records out
291050 bytes (291 kB, 284 KiB) copied, 0,136303 s, 2,1 MB/s
andrei@UBU1A:~$ ls -lR | dd bs=4096 of=ls.txt
71+1 records in
71+1 records out
291050 bytes (291 kB, 284 KiB) copied, 0,0373183 s, 7,8 MB/s
andrei@UBU1A:~$
```

```
andrei@UBU1A:~/labbl2/lab20$ ls
a.txt  b.txt  c.txt  wc.txt
andrei@UBU1A:~/labbl2/lab20$ diff a.txt b.txt
3,4c3,4
< e
< r
---
> ea
> rb
andrei@UBU1A:~/labbl2/lab20$ diff c.txt b.txt
1c1,6
< qwertyuiopasdfghjklzxcvbnm
---
> q
> w
> ea
> rb
> t
> yh
andrei@UBU1A:~/labbl2/lab20$ cp b.txt c.txt
andrei@UBU1A:~/labbl2/lab20$ diff c.txt b.txt
andrei@UBU1A:~/labbl2/lab20$
```

```
andrei@UBU1A:~/labbl2/lab20$ grep amo ls.txt
drwxrwxr-x  2 andrei andrei      4096 ноя 19 20:17 pryamoug
./pryamoug:
andrei@UBU1A:~/labbl2/lab20$
```

```
andrei@UBU1A:~/labbl2/lab20$ vim b.txt
andrei@UBU1A:~/labbl2/lab20$ join a.txt b.txt
q abc def
w def abc
t
yh
andrei@UBU1A:~/labbl2/lab20$
```

The image displays two side-by-side terminal windows. The left window shows the contents of 'a.txt', which is 6 lines long and 21 bytes in size. The right window shows the contents of 'b.txt', which is 6 lines long and 23 bytes in size. Both files contain the same text: 'q abc', 'w def', 'e', 'r', 't', and 'yh'.

```
q abc
w def
e
r
t
yh
```

"a.txt" 6L, 21B

```
q def
w abc
ea
rb
t
yh
```

"b.txt" 6L, 23B

SORT – вывод строк файла в лексикографическом порядке

```
andrei@UBU1A:~/labbl2/laba20$ sort -b -d a.txt
e
q
r
t
w
yh
```

TAIL – вывод послед 10 строк файла

```
andrei@UBU1A:~/labbl2/laba20$ tail ls.txt

./trash:
total 40
-rwxrwxr-x 1 andrei andrei 16240 янв 20 12:28 a
-rw-rw-r-- 1 andrei andrei 481 янв 20 12:35 a.c
-rwxrwxr-x 1 andrei andrei 16280 янв 20 16:49 b
-rw-rw-r-- 1 andrei andrei 498 янв 20 16:49 b.c

./Videos:
total 0
andrei@UBU1A:~/labbl2/laba20$
```

TEE – копирование stdin в stdout/файл.

```
andrei@UBU1A:~/labbl2/laba20$ ls
a.txt b.txt c.txt ls.txt wc.txt
andrei@UBU1A:~/labbl2/laba20$ tee d.txt
qefds
qefds
qfewdscsax^[[Z
qfewdscsax
ewfgthyuj
ewfgthyuj

^C
andrei@UBU1A:~/labbl2/laba20$
andrei@UBU1A:~/labbl2/laba20$ vim d.txt
andrei@UBU1A:~/labbl2/laba20$ more d.txt
qefds
qfewdscsax
ewfgthyuj

andrei@UBU1A:~/labbl2/laba20$
```

TR – замена символов из 1 набора на соотв символы из 2 набора (stdin -> stdout)

```
andrei@UBU1A:~/labbl2/laba20$ tr abcdef fedcba
a aa aaa ab ba bb bca ddd fff eof
f ff fff fe ef ee edf ccc aaa boa
```

UNIQ – вывод уникальных строк

```
andrei@UBU1A:~/labbl2/laba20$ uniq e.txt
a
ab
b
bb

andrei@UBU1A:~/labbl2/laba20$
```

```
a
a
a
a
a
a
a
a
ab
b
b
b
b
b
b
~
~
~
"e.txt" 14L, 29B
```

OD – побайтный вывод содержания файла

```
andrei@UBU1A:~/labbl2/laba20$ od e.txt
0000000 060412 060412 060412 060412 060412 060412 060412 060412
0000020 005142 005142 005142 005142 005142 061142 000012
0000035
```

SUM – вычисление контрольной суммы файла

```
andrei@UBU1A:~/labbl2/laba20$ sum e.txt
21293      1
andrei@UBU1A:~/labbl2/laba20$
```

CUT – вывод выбранных частей (каждой) строки файла

```
andrei@UBU1A:~/labbl2/laba20$ more cut.txt
qwerty
ytrewq
ewqerty
ytrqwe

andrei@UBU1A:~/labbl2/laba20$ more cut.txt | cut -b 1
q
y
e
y
```

VI/VIM – открытие одноименного текстового редактора.

```
andrei@UBU1A:~/labbl2/laba20$
andrei@UBU1A:~/labbl2/laba20$ vim vim.txt
```

```
"vim.txt" [New]
```

TAR – инструмент архивации (совокупность файлов архивируется в один файл)

```
andrei@UBU1A:~/labbl2/laba20$ tar --totals -cvf archive.tar e.txt
e.txt
Total bytes written: 10240 (10KiB, 6,0MiB/s)
andrei@UBU1A:~/labbl2/laba20$ ls
archive.tar a.txt b.txt c.txt cut.txt d.txt ed.txt e.txt
```

GZIP – архиватор (сжимает и архивирует)

```
andrei@UBU1A:~/labbl2/laba20$ vim g.txt
andrei@UBU1A:~/labbl2/laba20$ gzip g.txt
andrei@UBU1A:~/labbl2/laba20$ ls
a.txt c.txt d.txt f.txt ls.txt odd.txt tee.txt
b.txt cut.txt e.txt g.txt.gz oddd.txt od.txt wc.txt
andrei@UBU1A:~/labbl2/laba20$
```

HEAD – вывод верхней части файла (дефолтно 10 строк)

```
andrei@UBU1A:~/labbl2/laba20$ head ls.txt
.:
total 674924
-rw-rw-r-- 1 andrei andrei 0 окт 1 21:02 andrei@176.62.188.171
drwxrwxr-x 3 andrei andrei 4096 окт 1 21:09 andrei@176.62.188.171
drwxrwxr-x 4 andrei andrei 4096 окт 4 21:37 apache
-rw-rw-r-- 1 andrei andrei 0 фев 17 19:19 a.txt
-rw-rw-r-- 1 andrei andrei 197 окт 15 19:50 config.xml
drwxr-xr-x 3 andrei andrei 4096 сен 17 17:59 Desktop
drwxr-xr-x 2 andrei andrei 4096 сен 14 22:25 Documents
drwxr-xr-x 2 andrei andrei 4096 сен 14 22:25 Downloads
andrei@UBU1A:~/labbl2/laba20$
```

TOUCH – создать файл

```
andrei@UBU1A:~/labbl2/laba20/touch$
andrei@UBU1A:~/labbl2/laba20/touch$ ls
andrei@UBU1A:~/labbl2/laba20/touch$ touch touch.txt
andrei@UBU1A:~/labbl2/laba20/touch$ ls
touch.txt
```

Код скрипта:

```
#!/bin/bash
```

```
infile=$1
```

```
outfile=$2
```

```
#Проверка на наличие переменной входного и выходного файла.
```

```
if [[ ( -z $infile || -z "$outfile" ) ]];
```

```
then
```

```
    echo "Укажите имена входного и выходного файла.";
```

```
    exit;
```

```
fi
```

```
#Проверка на наличие входного файла в файловой системе.
```

```
if ! [ -f $infile ];
```

```
then
```

```
    echo "Входной файл $infile не существует.";
```

```
    exit;
```

```
fi
```

```
#Запись в выходной файл уникальных пятых колонок(подстрок) второго файла (с указанием частоты).
```

```
cat $infile | awk ' { print $5 } ' | sort | uniq -c | sort -n >> $outfile
```

7. Вывод:

В результате лабораторной работы были выполнены все поставленные цели. Изучены важные утилиты UNIX, использование которых будет полезно при автоматизации некоторых процессов по работе с ОС и её составляющими. Кроме этого, результатом лабораторной работы стал работающий скрипт, написание которого закрепило знания об утилитах.

Положительные стороны:

Написал первый скрипт в файле .sh, синтаксис имеет сходства уже изученными языками, вспомнил об уже забытые (grep) и открыл для себя новые (awk) полезные инструменты.

Отрицательные стороны:

Большой объем работы.