МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение высшего образования

«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ»

КАФЕДРА КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ

И ПРОГРАММНОЙ ИНЖЕНЕРИИ (КАФЕДРА №43)

ОТЧЕТ

ЗАЩИЩЕН С ОЦЕНКОЙ

ПРЕПОДАВАТЕЛЬ:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| старший преподаватель |  |  |  | М.Д. Поляк |
| должность, уч. степень, звание |  | подпись, дата |  | инициалы, фамилия |

ОТЧЕТ О ЛАБОРАТОРНОЙ РАБОТЕ №1

«РАБОТА С ТЕКСТОВЫМИ ПОТОКАМИ В КОМАНДНОМ ИНТЕРПРЕТАТОРЕ BASH»

ПО ДИСЦИПЛИНЕ «ОПЕРАЦИОННЫЕ СИСТЕМЫ»

РАБОТУ ВЫПОЛНИЛ

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| СТУДЕНТ ГР. | 4631 |  |  |  | С.А. Гришин |
|  |  |  | подпись, дата |  | инициалы, фамилия |

Санкт-Петербург

2019

**Цель работы** Изучение принципов работы с командным интерпретатором GNU/Linux и основ обработки текстовых файлов с помощью команд grep, awk, sed.

**Индивидуальное задание**

9. Написать скрипт с использованием grep, sed, awk (необходимо использовать не менее одной из указанных утилит; использовать все три необязательно) для переконвертирования данных в формат JSON со следующей структурой:

{

"Порядковый номер запроса": {

"timestamp": "Отметка времени, когда поступил запрос",

"client ip": "IP адрес пользователя",

"client port": "Порт пользователя"

},

"Порядковый номер запроса": {

"timestamp": "Отметка времени, когда поступил запрос",

"client ip": "IP адрес пользователя",

"client port": "Порт пользователя"

},

...

}

Взятый в кавычки русский текст необходимо заменить соответствующими значениями из логов. В формате JSON необходимо все ключи и все строковые значения помещать в двойные кавычки, числовые же значения (int, float) следует вставлять без кавычек. Для отступов использовать символ табуляции (4 пробела в примере выше = одна табуляция). Сохранить в файле results.json результат применения написанного скрипта к строкам с 51 по 100-ю (включительно) файла dns-tunneling.log.

**Описание входных данных**

3. Порядковый номер запроса: long,

4. Отметка времени, когда поступил запрос: два числа long, разделенных точкой; первое число – количество секунд, прошедших с 1 января 1970 года; второе число – количество микросекунд; т.е. фактически это тип данных float,

5. IP-адрес пользователя: character array,

6. Порт пользователя: int,

**Результат выполнения работы**

{

    "2103430730": {

        "timestamp": 1451001605.558599,

        "client ip": "10.116.74.176",

        "client port": 41280

    },

    "2103430748": {

        "timestamp": 1451001605.585656,

        "client ip": "10.47.144.132",

        "client port": 43301

    },

…

**Исходный код программы с комментариями**

#!/bin/bash

echo "{" > results.json

awk '

NR>=51 && NR<=100 {

printf "%s\t\"%s\": {\n\t\t\"timestamp\": %s,\n\t\t\"client ip\": \"%s\",\n\t\t\"client port\": %s\n\t}",

separator, $3, $4, $5, $6

separator = ",\n"

}' dns-tunneling.log >> results.json

printf "\n}\n" >> results.json

В начале идёт инициализация объекта в JSON, после вся работа передается в awk. На вход подается файл с логами и результат перенаправляется в results.json. В awk выбираются только строчки больше 51 и меньше 100 из входного файла. После этого идёт разбор каждой строки между 51 и 100 в переменные $3, $4, $5, $6. Эти переменные вставляются в заранее подготовленный JSON шаблон. В конце завершается JSON объект.

**Выводы**

Были получены знания по работе с обработкой текстовых данных с помощью команды awk. Часть файла с логами была преобразована в форматированный JSON файл.