|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования **«МИРЭА − Российский технологический университет»**  **РТУ МИРЭА** |

**Институт информационных технологий (ИИТ)**

**Кафедра прикладной математики (ПМ)**

**ОТЧЕТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ**

по дисциплине «Методы анализа данных»

**Практическая работа № 3**

|  |  |  |  |
| --- | --- | --- | --- |
| Студент группы ИНБО-01-17 | *ИМБО-01-21, Малкина В.В.* | (подпись) | |
| Преподаватель | *Буданцев А.В.* | (подпись) | |
| Отчет представлен | «\_\_\_»\_\_\_\_\_\_\_\_2023г. | |  | |

Москва 2023г.

Работа с apache flume

Получение данных с помощью Apache Flume

В этой лабораторной работе мы запустим агент Flume чтобы собрать данные из разных источников и сохранить их в HDFS или локальной файловой системе.

1. **Простой перенос данных**

Этот агент позволяет пользователю генерировать события и последовательно логгировать их в консоль. В следующей конфигурации определяется один агент под названием agent1.

Создадим файл конфигурации (Рисунок 1).

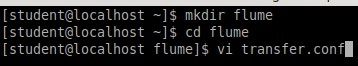


Рисунок 1 — Создание файла конфигурации

Зададим настройки файлу конфигурации агента Agent1 (Рисунок 2).

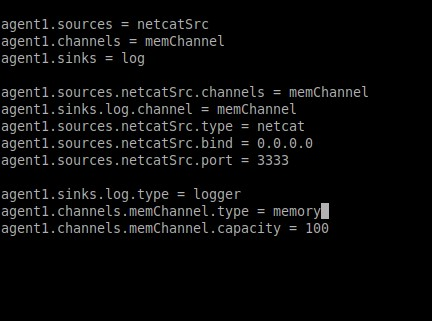


Рисунок 2 — Настройки файла конфигурации агента Agent1

В agent1 есть источник, который слушает порт 3333 для получения данных, канал, который буферизирует данные событий в памяти и слив, который логгирует данные событий в консоль.

Запустим Flume Agent1 (Рисунок 3).



Рисунок 3 — Запуск Agent1

Откроем ещё одно окно терминала и выполним команду telnet (Рисунок 4).

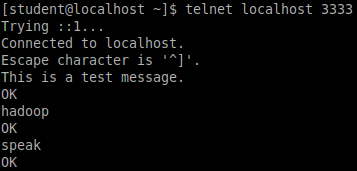


Рисунок 4 — Выполнение команды telnet

Проверим, что сообщение, отправленное в telnet, вывелось в терминале, где был запущен агент flume (Рисунок 5).

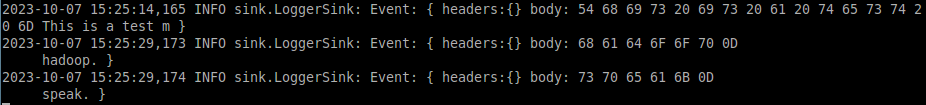


Рисунок 5 — Проверка работы команды

Закроем telnet (Рисунок 6).

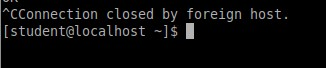


Рисунок 6 — Закрытие telnet

1. **Базовый перенос данных через буферную директорию**

Agent2 будет использоваться для сохранения файлов, попадающих в буферную директорию, в локальную директорию.

Создадим файл конфигурации (Рисунок 7).



Рисунок 7 — Создание файла конфигурации

Запишем настройки файла конфигурации Agent2 (Рисунок 8).

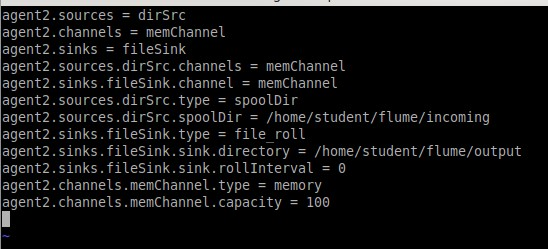


Рисунок 8 — Настройки файла конфигурации Agent2

Создадим все необходимые папки (Рисунок 9).

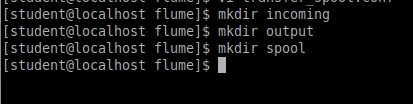


Рисунок 9 — Создание папок

Запустим Agent2 (Рисунок 10).



Рисунок 10 — Запуск Agent2

Откроем еще одно окно терминала и скопируем несколько файлов в буферную директорию (Рисунок 11).

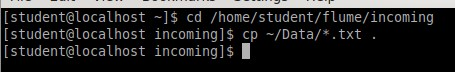


Рисунок 11 — Копирование txt файлов

Создадим файл hello.txt и запишем туда текст (Рисунок 12).

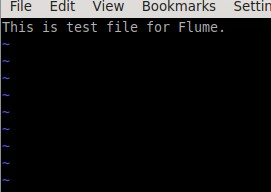


Рисунок 12 — Создание файлов hello.txt

Проверим выполнение копирования (Рисунок 13).

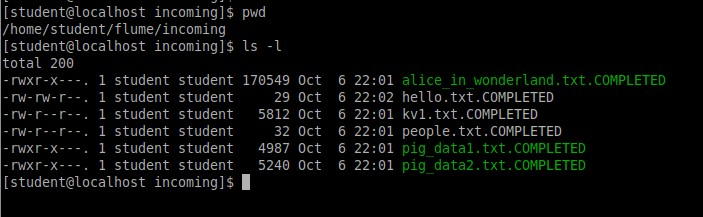


Рисунок 13 — Содержимое директории incoming

Мы должны увидеть сообщение о том, что pig\_data1.txt, pig\_data2.txt, alice\_in\_wonderland.txt и hello.txt скопированы в буферную директорию. Сообщение появится в терминале, где запущен Agent2 (Рисунок 14).

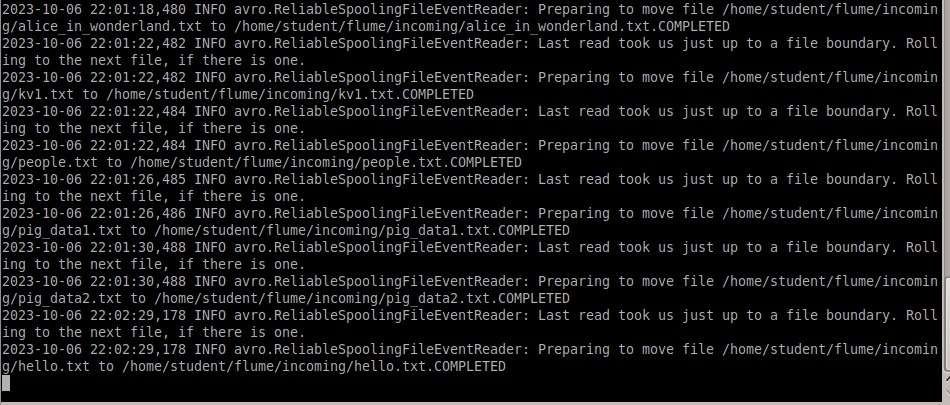


Рисунок 14 — Вывод терминала

Перенесенные файлы сохранены в виде файлов в директории output (Рисунок 15).

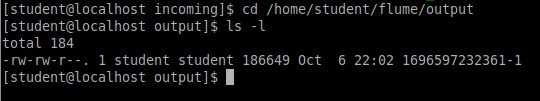


Рисунок 15 — Содержимое директории output

1. **Использование Перехватчика (Interceptor)**

Agent3 будет использоваться для вставки IP адреса хоста, где запущен агент, в заголовки события.

Создадим файл конфигурации (Рисунок 16).



Рисунок 16 — Создание файла конфигурации

Запишем настройки для файла конфигурации Agent3 (Рисунок 17).

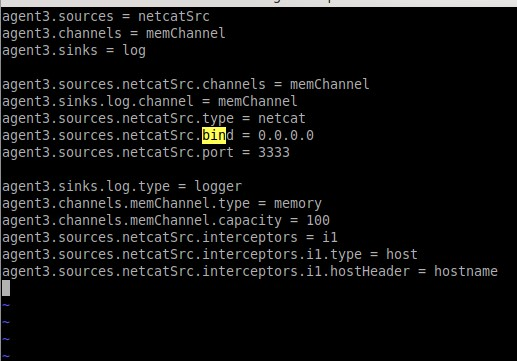


Рисунок 17 — Настройки файла конфигурации Agent3

Запустим Agent3 (Рисунок 18).



Рисунок 18 — Запуск Agent3

Откроем ещё одно окно терминала и выполним команду telnet (Рисунок 19).

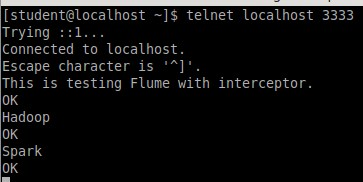


Рисунок 19 — Запуск команды telnet

Сообщение, отправленное в telnet, выводится с терминала, где агент был запущен, и подтверждается, что IP-адрес хоста, на котором в данный момент запущен агент, вставляется в заголовок события и передается дальше (Рисунок 20).

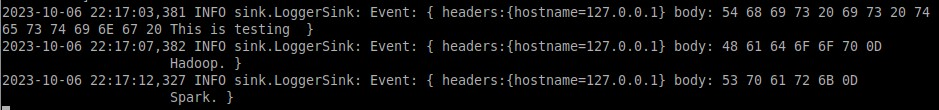


Рисунок 20 — Вывод терминала

Удалим временную директорию, использовавшуюся для операций Flume (Рисунок 21).



Рисунок 21 — Удаление временной директории

1. **Создайте новый поток данных «с нуля»**

Создадим файл конфигурации (Рисунок 22).

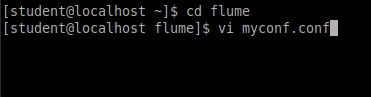


Рисунок 22 — Создание файла конфигурации

Запишем настройки для файла конфигурации Agent4 (Рисунок 23).

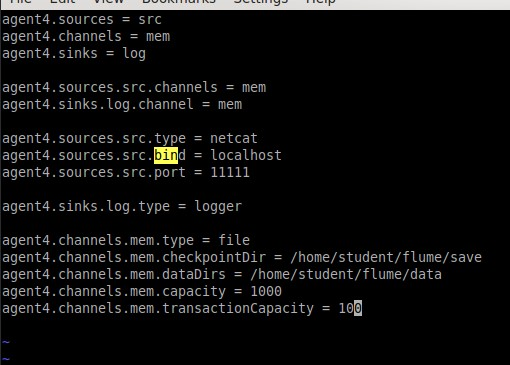


Рисунок 23 — Настройки файла конфигурации Agent4

Создадим необходимые все необходимые папки (Рисунок 24).



Рисунок 24 — Создание необходимых папок

Запустим Agent4 (Рисунок 25).



Рисунок 25 — Запуск Agent4

В другом окне терминала запустим telnet и подключимся к порту 11111. Напишем что-нибудь и убедимся, что мы видим написанное в терминале, где запущен агент (Рисунок 26).

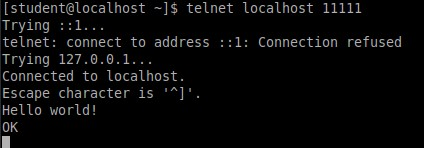


Рисунок 26 — Запуск telnet

Проверим вывод в терминале (Рисунок 27).

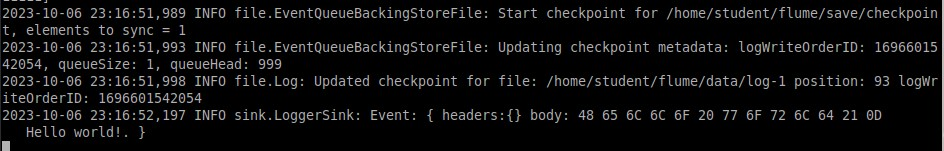


Рисунок 27 — Вывод терминала

Проверим, что в директории data создались новые файлы (Рисунок 28).

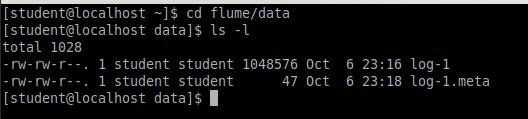


Рисунок 28 — Содержимое директории data