|  |
| --- |
| 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 № 5**

|  |  |  |  |
| --- | --- | --- | --- |
| 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 Kafka

Создание темы, производителя и потребителя Kafka

В этой работе мы будем использовать сервисы Kafka в командной строке для создания темы. Мы также будем использовать Kafka для создания производителей и потребителей для передачи данных через них.

1. **Настройка Apache kafka**

Остановим сервисы HBase (Рисунок 1).

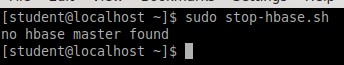


Рисунок 1 — Остановка HBase

Перезапустим Kafka и Zookeeper, убедимся, что они запущены (Рисунок 2).

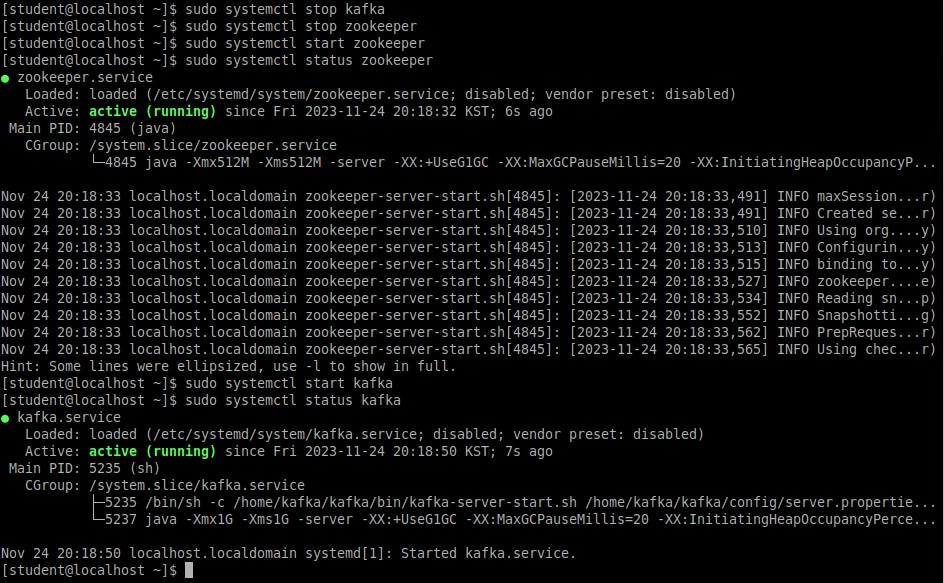


Рисунок 2 — Запуск zookeeper и kafka

1. **создание темы kafka**

Напишем необходимые команды для создания темы (Рисунок 3).

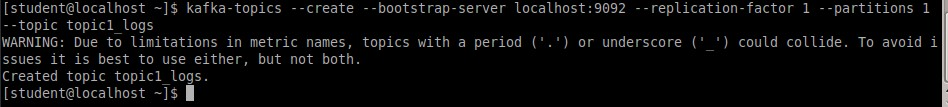


Рисунок 3 — Создание темы

Отобразим список тем Kafka (Рисунок 4).



Рисунок 4 — Список тем

Посмотрим тему topic1\_logs (Рисунок 5).

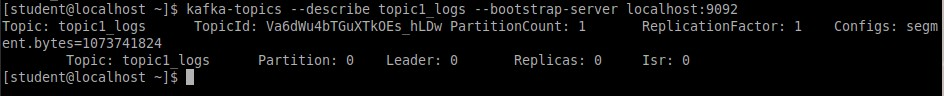


Рисунок 5 — topic1\_logs

1. **Создание потребителей и производителей для темы**

Откроем 2 окна терминала и создайте производителя в одном окне, а потребителя в другом.

Из первого терминала запустите команду kafka-console-producer для запуска производителя (Рисунок 6).



Рисунок 6 — Запуск производителя

Из второго терминала вызовите команду kafka-console-consumer для создания потребителя (Рисунок 7).

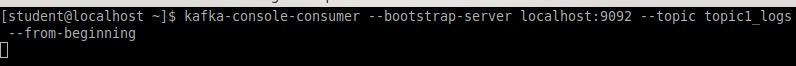


Рисунок 7 — Запуск потребителя

Переименуем терминалы (Рисунок 8).

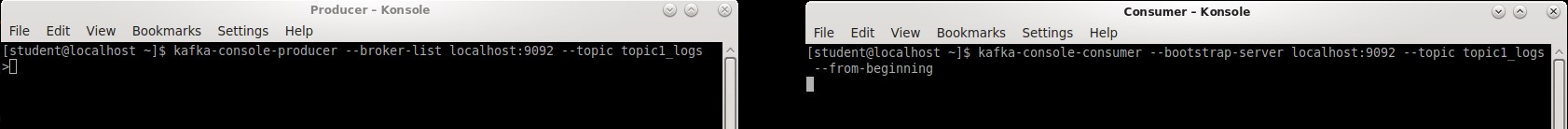


Рисунок 8 — Переименование терминалов

Начнем писать сообщения в Producer и увидим их отображение в Consumer (Рисунок 9).

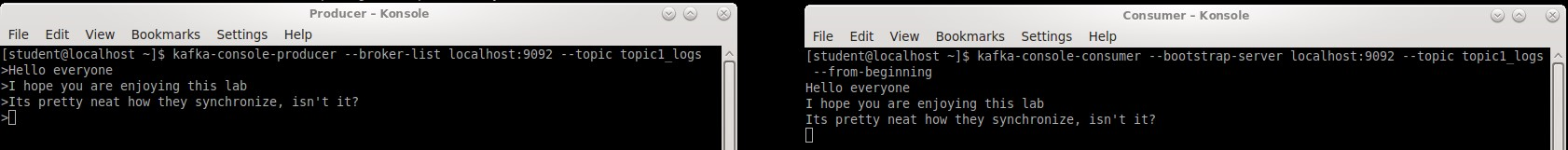


Рисунок 9 — Создание сообщений

Отправим содержимое файла Alice-in-Wonderlands.txt в тему topic\_logs (Рисунок 10).

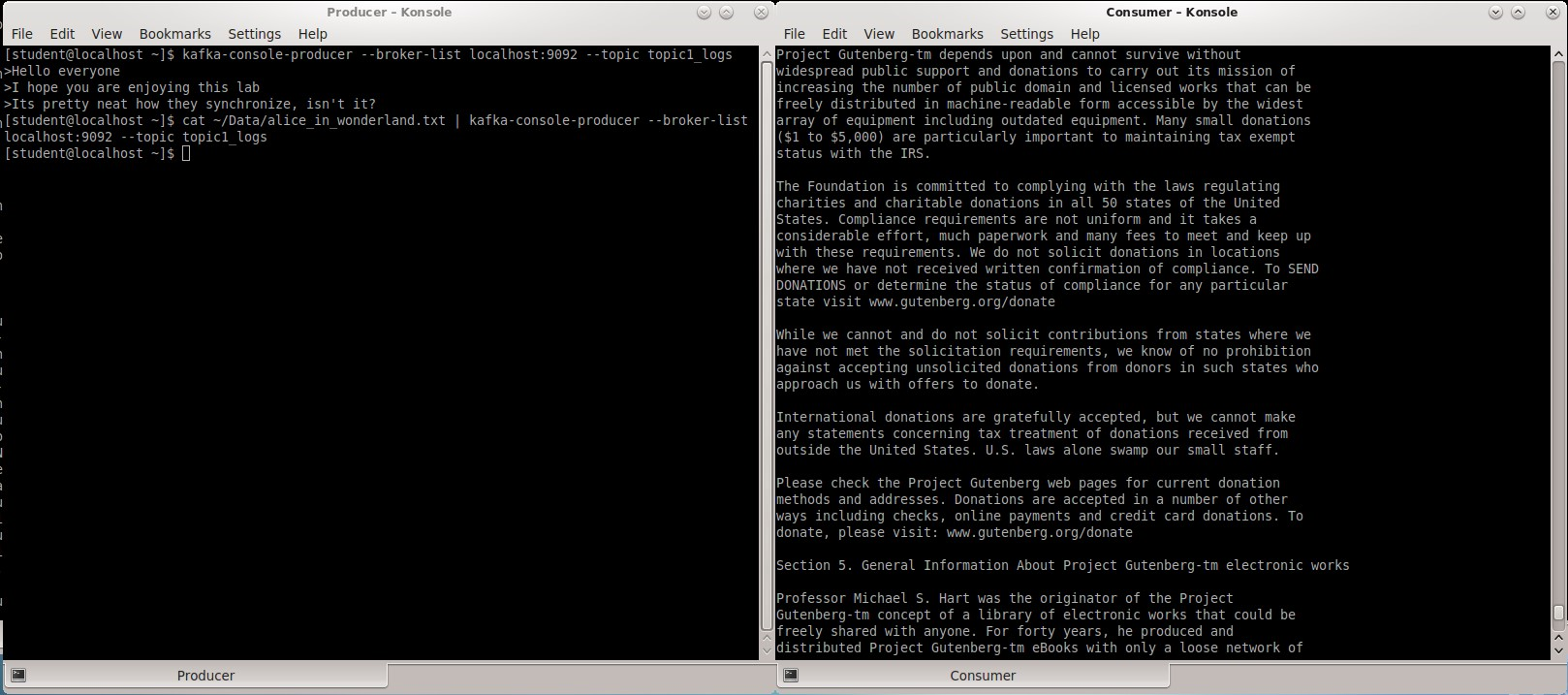


Рисунок 10 — Отправка файла Alice-in-Wonderlands

Отправка сообщений из Flume в Kafka

1. **настройка apache kafka**

Остановим HBase (Рисунок 11).

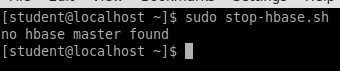


Рисунок 11 — Остановка HBase

Перезапустим Kafka и Zookeeper (Рисунок 12).

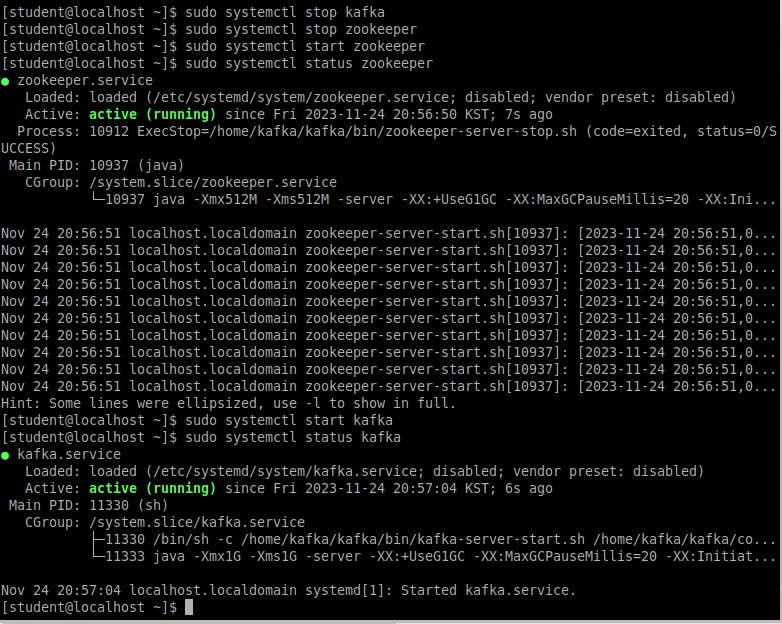


Рисунок 12 — Перезапуск kafka и zookeeper

1. **Создание конвейера Flume для сбора текста в потоковом режиме**

Перейдем в папку /home/student/Labs/C3U4 в терминале, а затем создадим новую папку spooldir и файл конфигурации (Рисунок 13).

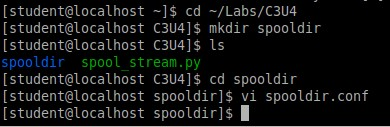


Рисунок 13 — Создание папки и файла конфигурации

Пропишем необходимые настройки:

* установим в качестве spool-директории /home/student/Labs/C3U4/spool;
* создадим 2 канала типа memory;
* настроим емкость (capacity) на 10000;
* настроим максимальное число событий на транзакцию (transactionCapacity) на 100;
* создадим слив Kafka и Logger;
* установим тип слива на org.apache.flume.sink.kafka.KafkaSink;
* установим тему Kafka в сливе на stream\_text;
* установим список серверов начальной загрузки (bootstrap) на localhost:9092;
* установим количество обрабатываемых сообщений в пакете (flumeBatchSize) на 5;
* установим тип на logger;
* подсоединим источник к обоим каналам и к каждому каналу по сливу.

Покажем содержимое (Рисунок 14).

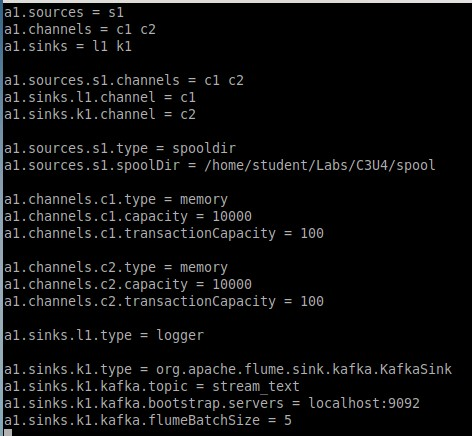


Рисунок 14 — Файл конфигурации

1. **Создание консольного потребителя kafka**

Откроем новый терминал и назовите его «Kafka Consumer», запустим консольного потребителя Kafka (Рисунок 15).

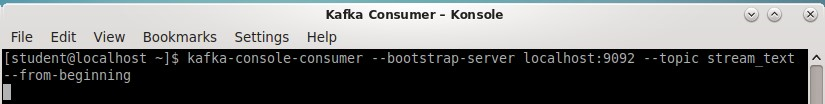


Рисунок 15 — Запуск консольного потребителя

1. **тестирование конвейера данных**

Перейдем в /home/student/Labs/C3U4, создадим каталог и назовите его spool (Рисунок 16).

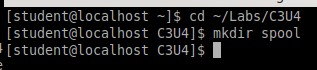


Рисунок 16 — Создание каталога spool

Откроем новый терминал и изменим имя терминала на «Kafka Producer», запустим поток данных Flume (Рисунок 17).

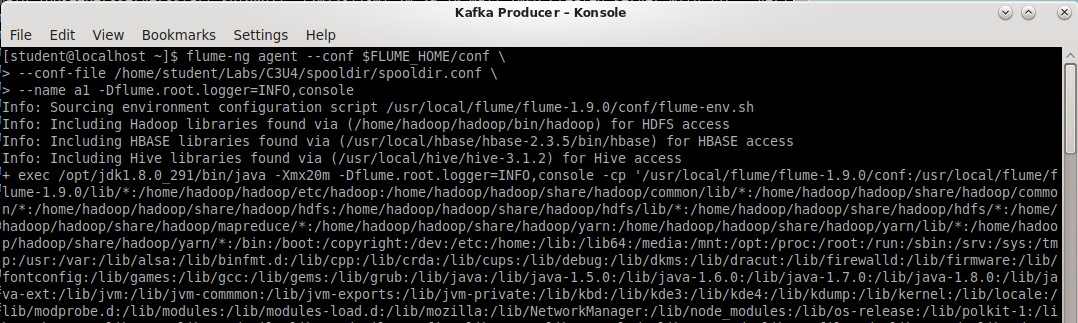


Рисунок 17 — Запуск потока данных Flume

Откроем новый терминал и назовем его «Streaming Source», перейдем в /home/student/Labs/C3U4 и выполним программу spool\_stream.py (Рисунок 18).

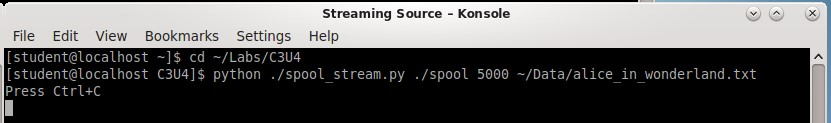


Рисунок 18 — Запуск программы spool\_stream

Посмотрим выходные данные с терминала Producer и Consumer (Рисунок 19).



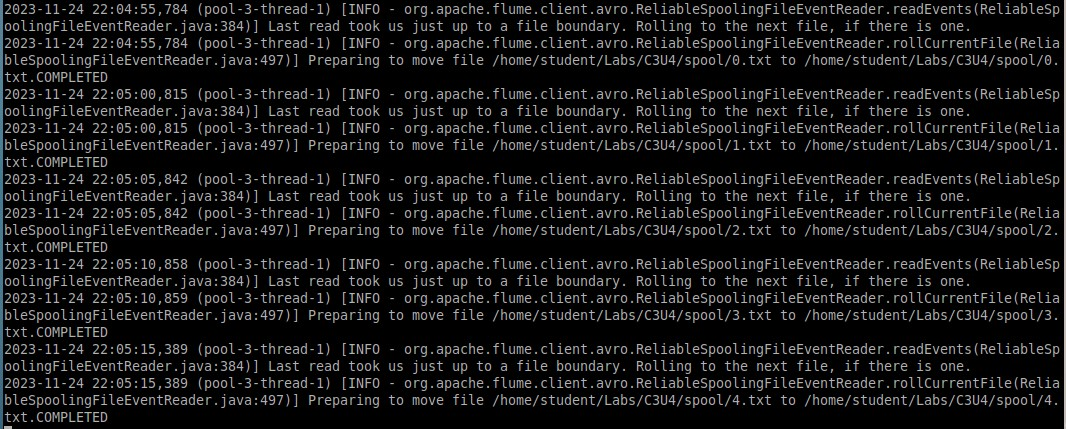


Рисунок 19 — Выходные данные терминалов