Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Кафедра автоматизації проектування

 енергетичних процесів і систем

**Звіт**

**з циклу лабораторних робіт**

**з дисципліни**

**„Обробка потокової інформації”**

Лабораторна  робота № 6

Організація спільної роботи Apache Kafka і Apache Spark Streaming в середовищі R.

Виконала: **Юрченко Богдана**

Студентка групи:   **ТВ-01 мп**

Перевірив викладач: **Іван Матичин**

Київ - 2021

Завдання:

1. Spark читає дані з topic1 за допомогою функції stream\_read\_kafka пакету

sparklyr.

2. Виконує деяку примітивну обробку/аналіз.

3. Надсилає оброблені дані у вигляді потоку до topic2 у Kafka за допомогою функції

stream\_write\_kafka пакету sparklyr.

В рамках даної роботи було зроблено :

змінено код консьюмера, продьсера та коду генерації топіків із попердніх робіт.

Кроки виконання :

1. Були запущені контейнери із zookeeper i kafka за допомогою докеру та докер-компоуз (аналогічно до першої роботи).

Код створення топіків 1 та 2 :



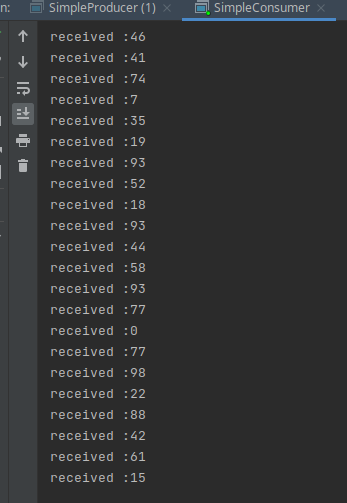
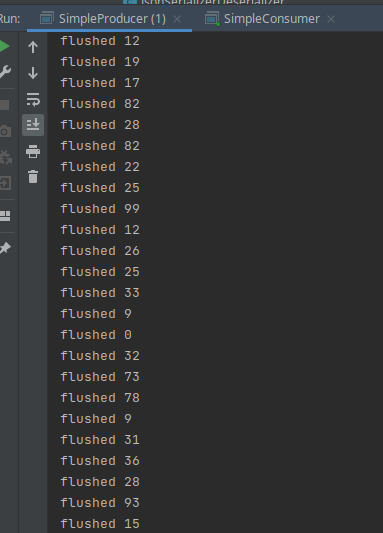
Перероблений продюсер, що відсилає в топік із числами в **топік 1** :



Код консьюмера, що читає числа із **топіку 2 :**

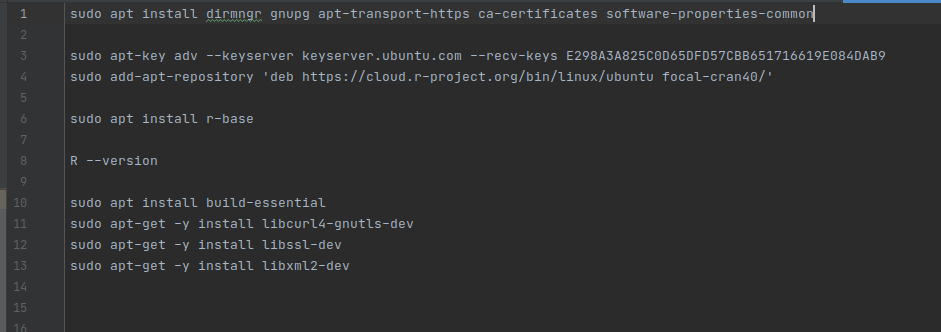


Результат роботи продьюсера та консьюмера

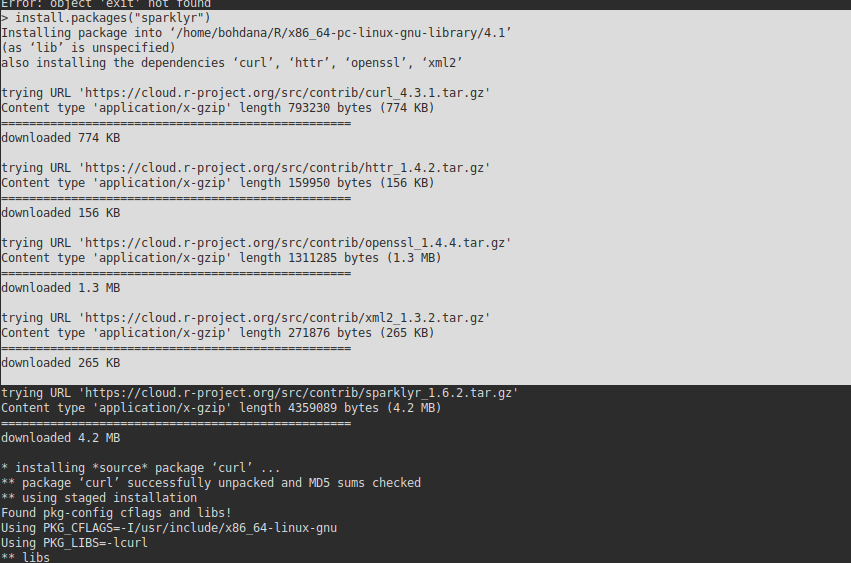


Між топіками 1 та 2 був створений додатковий шар із використанням Spark та R.

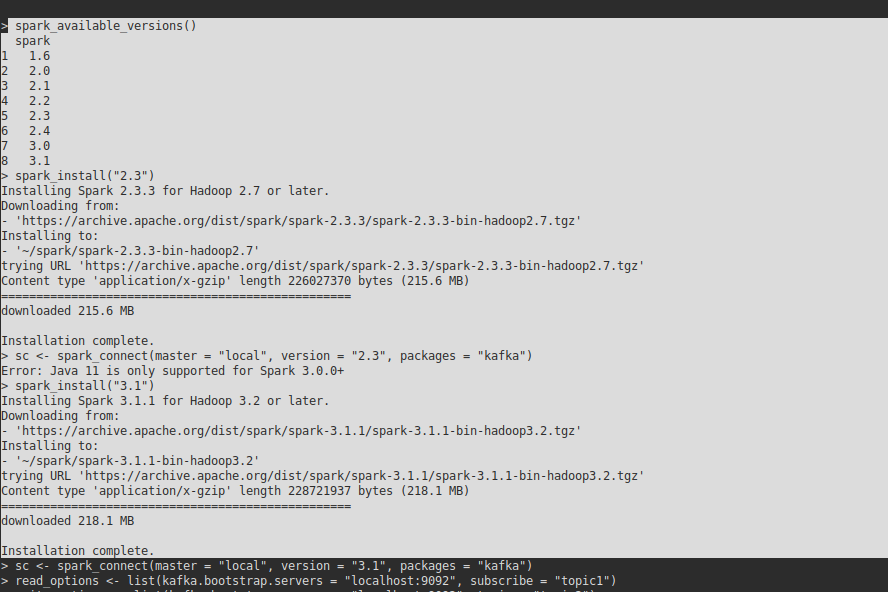
Для цього було встановлене дев середовище для R за допомогою команд :



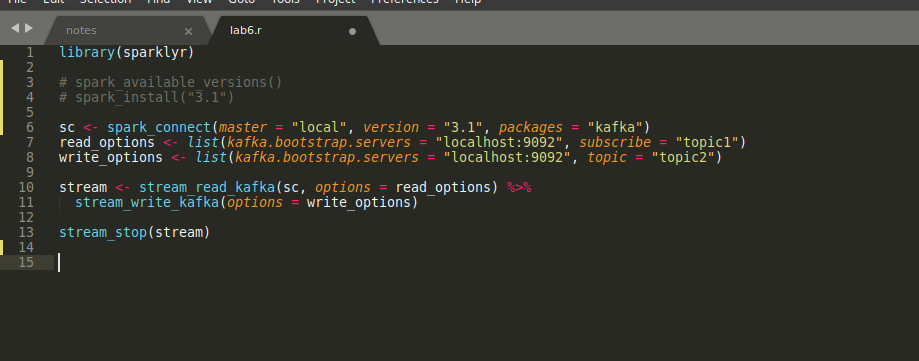
Після цього уже в середовищі R був встановлений пакет Sparklyr :



Після чого було встанолено дев версію спарка локально за допомогою команд :



Після цьго покрокове виконання команд із файлу зстворило стрім днних, що перенаправляв числа із топіку 1 в топік 2 кафки :



Лише після цього дані почали поступати до консьюмера.

Закінчуємо роботу :

