Федеральное агентство связи

Федеральное государственное бюджетное образовательное учреждение высшего образования

«Сибирский государственный университет телекоммуникаций и информатики»

(СибГУТИ)

Кафедра вычислительных систем

**Лабораторная работа №3**

«Анализ графа пользователей Twitter»

Выполнил:

студент гр. МГ-211 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ / Бурдуковский И.А./

подпись

Проверил:

Профессор

кафедры ВС \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ /Курносов М.Г./

ОЦЕНКА, подпись

Оглавление

[Задание 3](#_Toc133933349)

[Описание работы 4](#_Toc133933350)

# Задание

Граф пользователей с информацией о их подписчиках хранится в текстовом файле, каждая строка которого имеет следующий формат:

<user\_id> <follower\_id>

Файл размещен на кластере:

/home/ipa/pub/hadoop/lab34/followers.db

Требуется:

1. Вычислить распределение количества подписчиков (статистическое распределение выборки)
2. Определить Top50 пользователей по числу подписчиков
3. Вычислить среднее количество подписчиков

Программы должны быть написаны на языке Java (Apache Hadoop Java API).

# Описание работы

Первым делом был реализован подсчёт количества подписчиков у каждого пользователя. Для быстрой сборки и запуска приложения были реализованы скрипты.

В результате на HDFS будет создан каталог ./lab3/countfollower-output, куда программа CountFollowers.java сохранит полученный после обработки набор данных в формате:

<user\_id> <followers\_count>

Данный набор будет использоваться в дальнейших вычислениях.

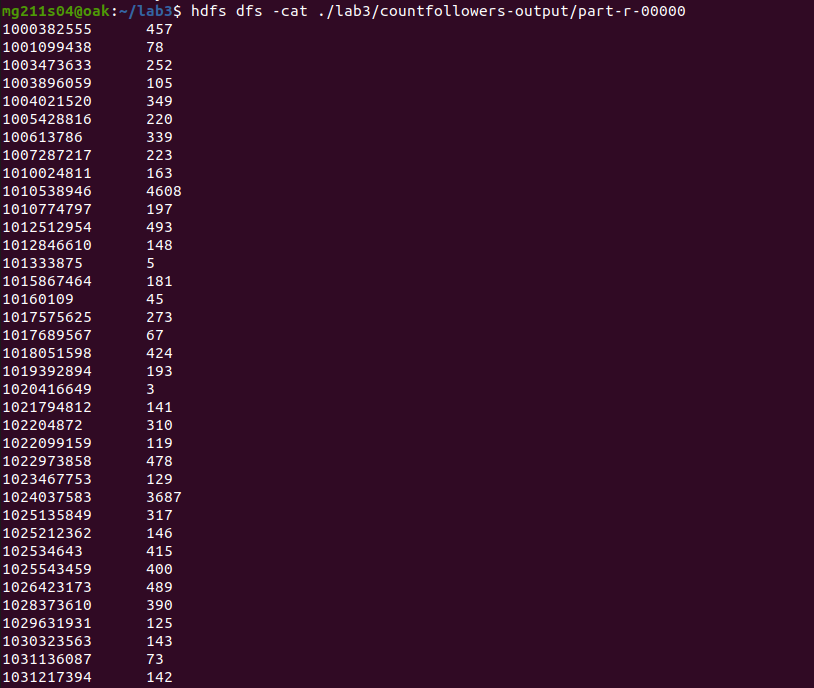
Запуск сборки:



Запуск приложения:



Результат:



1. Для вычисления распределения количества подписчиков был реализован DistributionCountFollowers.java, который сохранит полученные данные в каталог HDFS ./lab3/distribution-count-followers-output в формате:

<followers\_count> <users\_count>

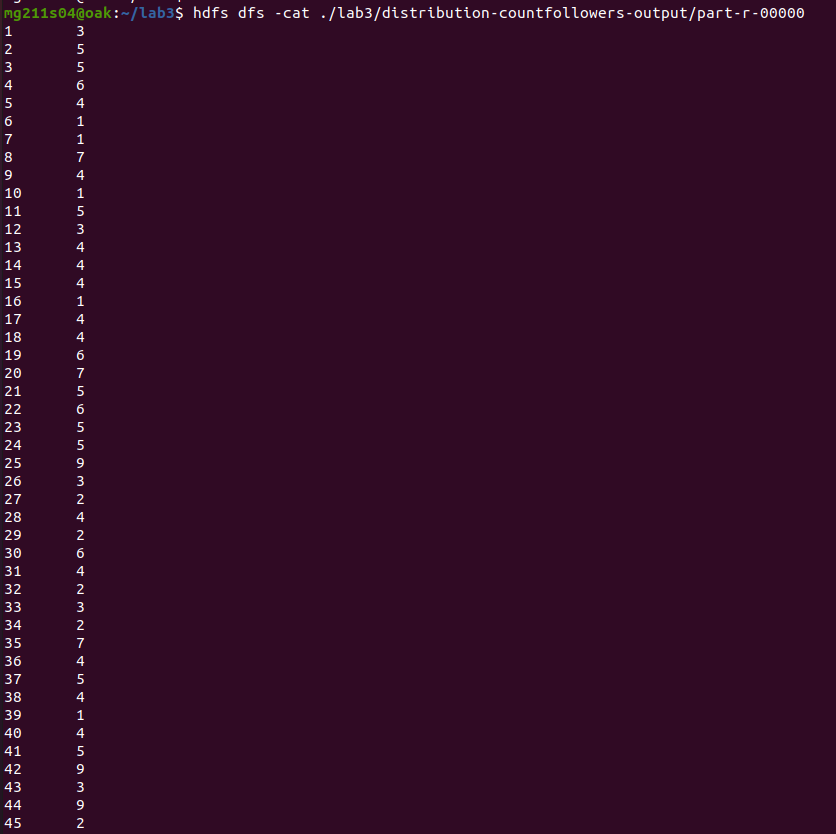
Запуск сборки:



Запуск приложения:

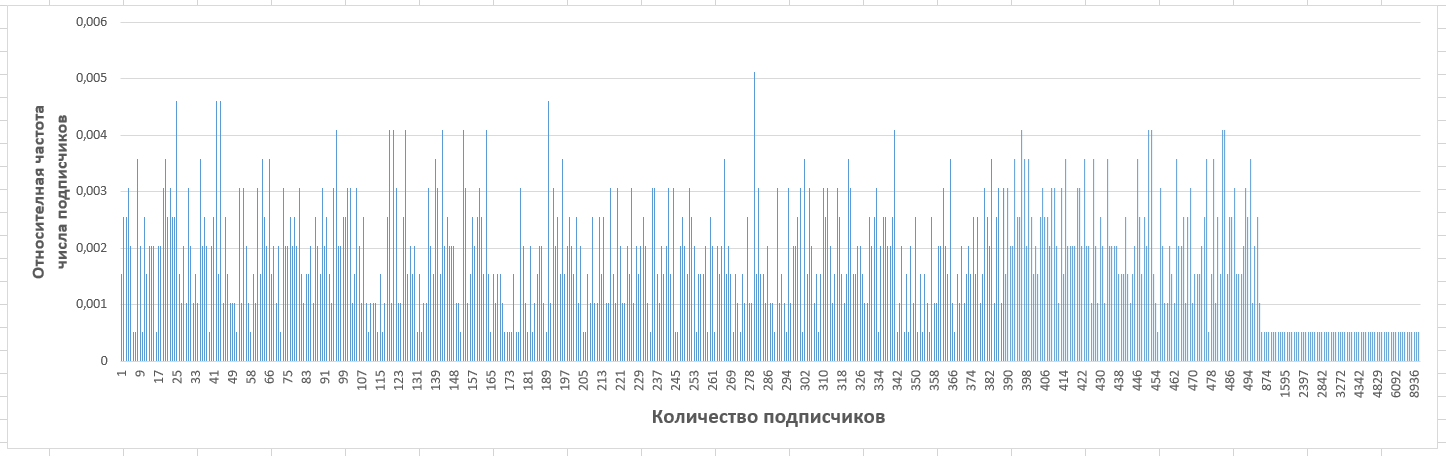


Результат:



Диаграммы распределения





1. Для определения TOP-50 пользователей по числу подписчиков был реализован Top50Users.java.  
   Каталог на HDFS с результатом: ./lab3/topusers-output

Формат: <user\_id > < followers\_count >

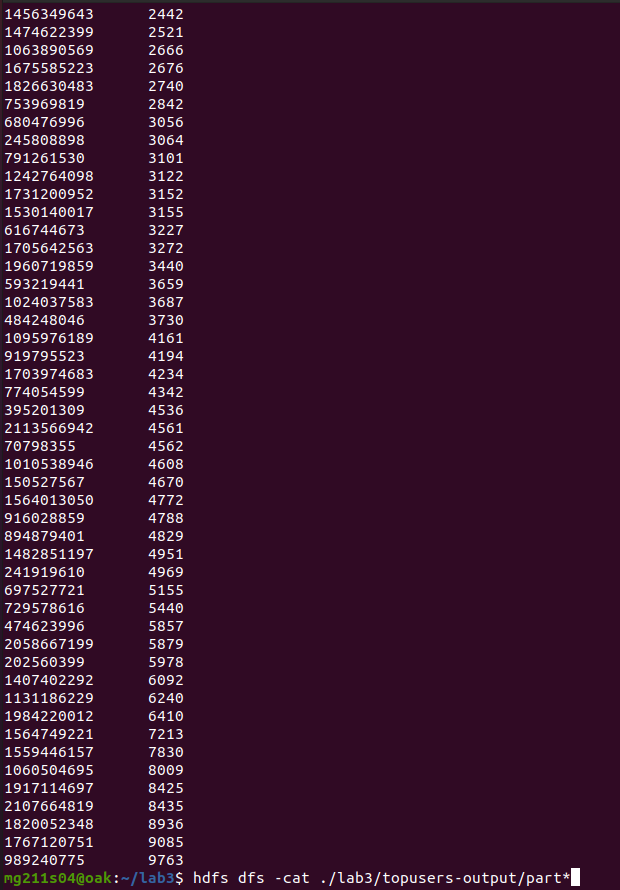
Запуск сборки:



Запуск приложения:



Результат:



1. Для вычисления среднего количества пользователей по числу подписчиков был реализован AvgCountFollowers.java.  
   Каталог на HDFS с результатом: ./lab3/avgcountfollowers-output

Запуск сборки:



Запуск приложения:



Результат:

