Отчёт

Студент: Турко С.А.
Группа: М16-522

Link: https://github.com/sergeyturko/DSBDA_HW1

Задача.

Необходимо найти самое длинное слово в .txt файле. Формат выходного файла – CSV.

Реализация.

Реализация задачи осуществлялась на ОС Windows 7x64. Использовался hadoop 2.6.5, установлен в соответствии с инструкцией http://www.ics.uci.edu/~shantas/Install_Hadoop-2.6.0_on_Windows10.pdf

Для сборки проекта использовалась IDE agnostic build Maven. pom.xml файл находится в репозитории. Для создания файловой системы HDFS и копирования данных был написан .bat файл hdfs.bat (в директории scripts). Скриншоты результатов представленны ниже.

Сборка производиться из командной строки командой <mvn package>

Для тестов использовались файлы testLine1.txt и testLine2.txt

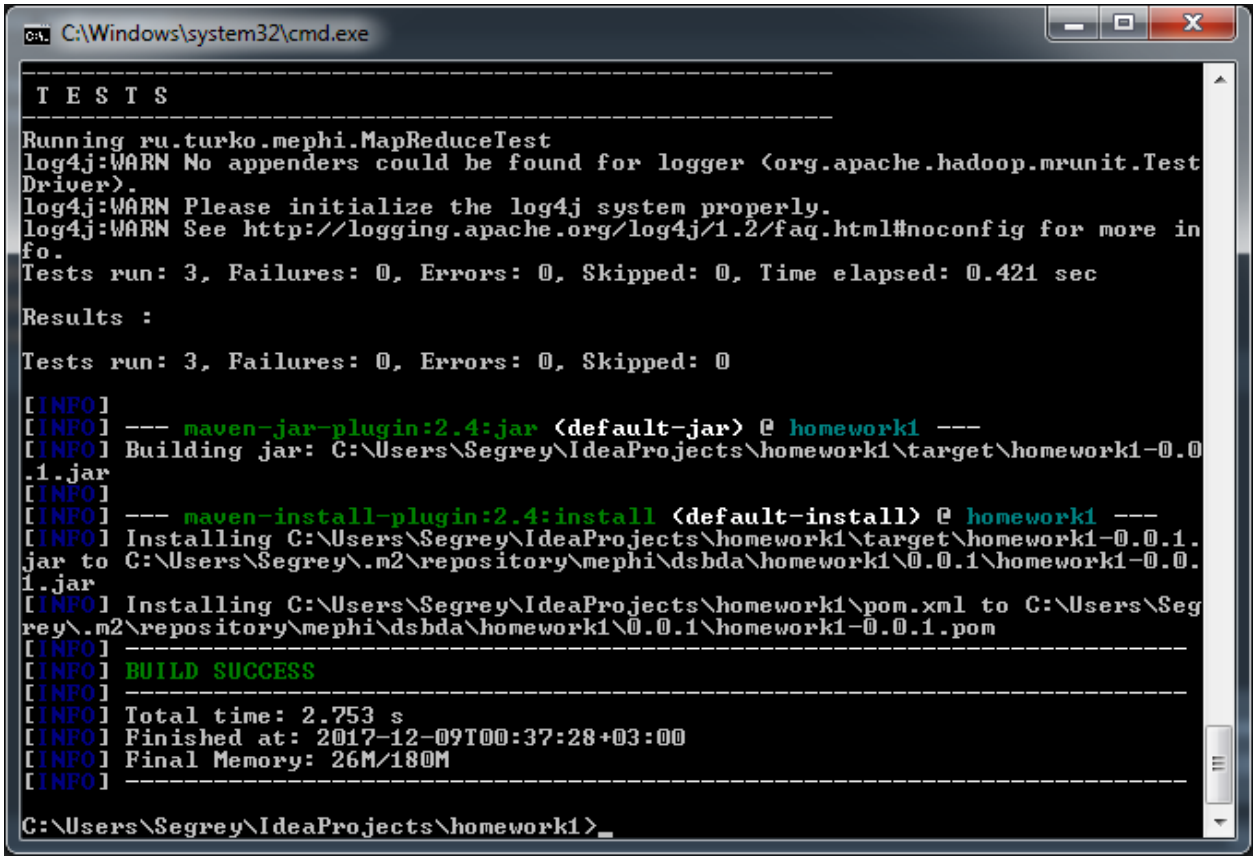

Рисунок 1 – Успешное выполнение тестов.

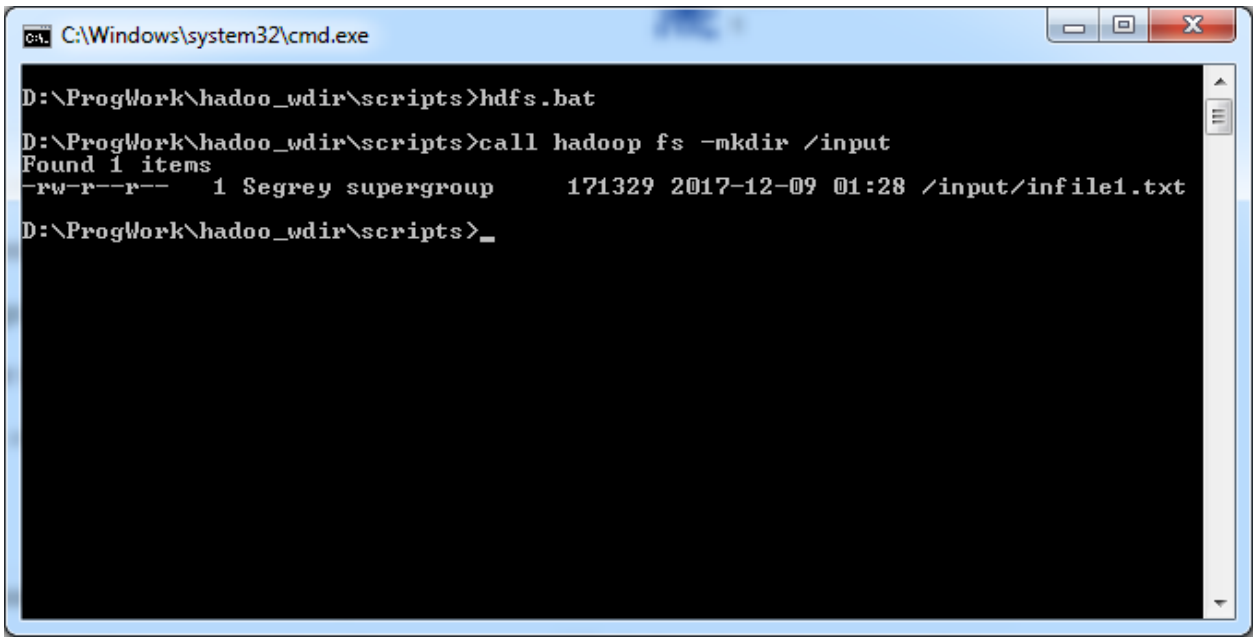

Рисунок 2 – Запуск скрипта. (Копирование файла в HDFS)

Запуск производиться командой:

hadoop jar homework1-0.0.1.jar homework1 /input /output

Рисунок 3 – Результат выполнения.