Ковальчук Александр, 520 группа
Задание №4 по курсу **"Современные методы распределенного хранения и обработки данных"**

Результаты запусков представлены в следующих таблицах:

2 узла A3

Количество Reducer`ов	Время выполнения (утилита time)	Время выполнения (Счетчики Hadoop ms)	Объем переданных данных (байт)
1	6m0.958s	1512478	1446996831
2	5m10.261s	1551230	1094370857
4	5m23.933s	1788784	1095381049
6	5m57.497s	2112796	1096394999
8	5m37.354s	1949706	1097410493

4 узла A3

Количество Reducer`ов	Время выполнения (утилита time)	Время выполнения (Счетчики Hadoop ms)	Объем переданных данных (байт)
1	5m29.548s	1659905	1446984274
2	4m16.299s	1396202	1094358229
4	4m15.298s	1705149	1095370567
6	4m8.983s	1672594	1096384393
8	4m2.197s	1813812	1097399747
10	4m16.747s	1930960	1098416777
12	4m26.269s	2123652	1099435173
14	4m34.114s	2191125	1100454737
16	4m50.047s	2370007	1101476109

В качестве счетчиков Hadoop считалось время «Total time spent by all map tasks» + «Total time spent by all reduce tasks». В качестве переданной информации считалась сумма счетчиков «<FILE, WASB>: Number of bytes <read, written>».

Как можно увидеть из таблиц:

- Объем переданной информации увеличивается с ростом числа редьюсеров. Это происходит, поскольку большему числу редьюсеров приходится передавать больше служебной информации
- Счетчики Hadoop учитывают суммарное время работы всех задач, поскольку данное время существенно отличается от времени, полученного с помощью утилиты time (данное время будет говорить, когда закончилось выполнения задачи пользователя, поэтому его можно считать репрезентативным)
- Самое быстрое время выполнения для двух узлов получилось при 4 редьюсерах
- Самое быстрое время выполнения для четырех узлов получилось при 8 редьюсерах