Ковальчук Александр, 520 группа
Задание №3 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	6m17.106s	858 870	3 796 630 581
2	6m1.951s	1 086 180	3 793 325 683
4	5m25.768s	1 445 693	3 793 979 907
6	5m36.702s	1 748 113	3 794 739 479
8	5m40.948s	1 661 059	3 795 773 851

4 узла A3

Количество Reducer`ов	Время выполнения (утилита time)	Время выполнения (Счетчики Hadoop ms)	Объем переданных данных (байт)
1	6m40.061s	867 550	3 796 630 581
2	5m41.696s	1 015 503	3 793 325 683
4	5m18.116s	1 411 853	3 793 979 907
6	5m25.432s	1 908 894	3 794 739 479
8	5m7.184s	2 183 366	3 795 773 851
10	5m10.262s	2 556 032	3 796 070 888
12	5m0.540s	2 781 293	3 797 355 394
14	5m26.253s	2 965 127	3 797 637 811
16	5m18.389s	2 735 454	3 799 162 963

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 получилось при количестве редьюсеров, кратном числу рабочих узлов (на каждый узел назначается одинаковое количество редьюсеров)