

Ковальчук Александр, 520 группа

Задание по курсу "Современные методы распределенного хранения и обработки данных" №2

	2 узла A3			4 узла A3		
	Время		Переданная информация (байт)	Время		Переданная информация (байт)
	Утилита time	Счетчики Hadoop (ms)		Утилита time	Счетчики Hadoop (ms)	
1 редьюсер	2m20.587s	226 092	1 164 569 076	2m23.248s	227 943	1 164 569 157
2 редьюсера	2m21.399s	232 971	1 164 694 288	2m21.138s	235 925	1 164 694 384
3 редьюсера	2m25.544s	244 755	1 164 819 363	2m17.531s	236 163	1 164 819 474
4 редьюсера	2m22.774s	252 276	1 164 944 438	2m19.203s	244 547	1 164 944 564
5 редьюсеров	-	-	-	2m18.814s	248 662	1 165 069 654
6 редьюсеров	2m28.000s	275 513	1 165 194 588	2m24.121s	262 252	1 165 194 744
7 редьюсеров	-	-	-	2m20.957s	267 801	1 165 319 834
8 редьюсеров	2m35.248s	289 140	1 165 444 738	2m19.793s	271 304	1 165 444 924
12 редьюсеров	-	-	-	2m23.688s	326 815	1 165 945 299
16 редьюсеров	-	-	-	2m33.874s	349 734	1 166 445 667

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 таблицы выше:

- Объем переданной информации увеличивается с ростом числа редьюсеров. Эта закономерность логична, поскольку требуется передавать больше данных большему числу Reducer'ов
- Счетчики Hadoop учитывают суммарное время работы всех Map и Reduce задач, поскольку фактически данное время больше времени, полученного с помощью утилиты time
- Варьирование малого количества редьюсеров не сильно влияет на время выполнения, однако с их увеличением время выполнения растёт