《分布式编程模型与系统》实验报告

报告题目：分布式计算系统——不同 Shuffle 类型的比较

姓　　名：陈越

学　　号：10195501438

完成日期：2022 年 6 月 24 日星期五

华东师范大学数据科学与工程学院实验报告

课程名称：分布式编程模型与系统　　**年级**：2019　　**上机实践成绩**：

指导教师：徐辰　　**姓名**：陈越　　**学号**：10195501438

上机实践名称：分布式编程模型与系统期末考查作业　　**上机实践日期**：6 月

分布式编程模型与系统——Shuffle 类型比较

一、 实验背景

Spark Shuffle 主要分为两种——基于 Hash 的 Shuffle 和基于 Sort 的 Shuffle。

Spark 0.8 以前：默认使用 Hash-Based Shuffle；

Spark 0.8：为 Hash-Based Shuffle 引入 File Consolidation 机制；

Spark 1.1：引入 Sort Based Shuffle；

Spark 1.2：默认 Shuffle 方式改为 Sort-Based Shuffle；

Spark 1.4：引入 Tungsten Sort-Based Shuffle；

Spark 1.6：Tungsten Sort-Based Shuffle 并入 Sort-Based Shuffle；

Spark 2.0：Hash-Based Shuffle 废弃；

为什么 Spark 放弃了 Hash-Based Shuffle，使用了 Sorted-Based Shuffle？

二、 实验目的

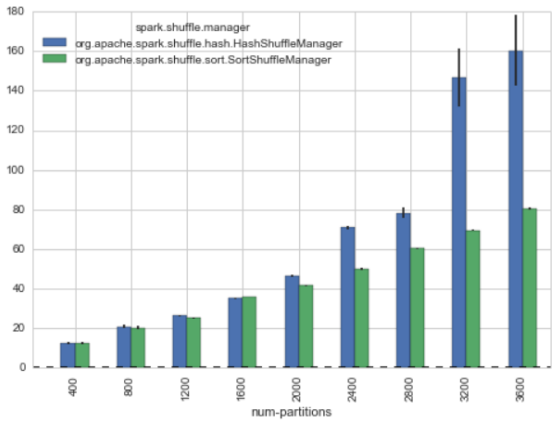
比较 Spark 中不同 shuffle 类型对程序执行性能的影响；

- Hash-Based Shuffle
- Sort-Based Shuffle

三、 相关工作

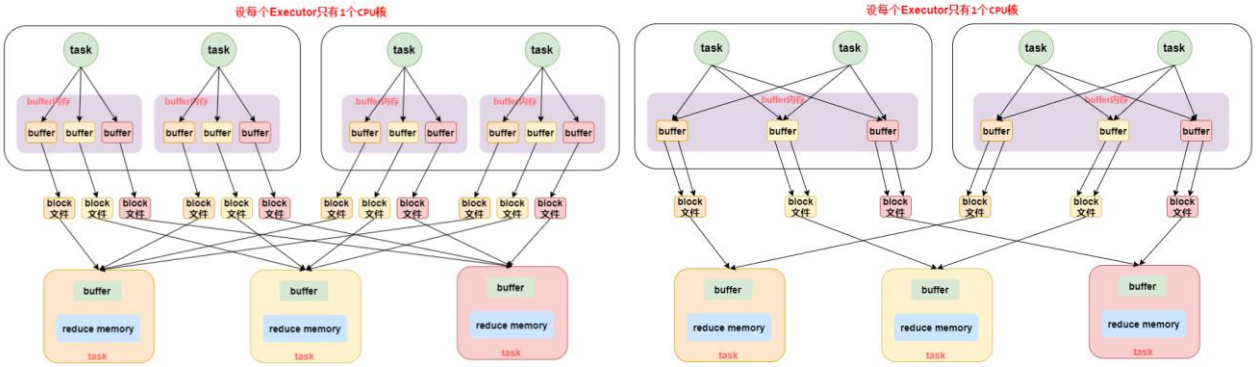
经查阅相关工作，相关实验结果如下(HashShuffle 和 SortShuffle 的执行时间比较)：

				spark.shuffle.manager	org.apache.spark.shuffle.hash.HashShuffleManager	org.apache.spark.shuffle.sort.SortShuffl
num-partitions	reduce-tasks	num-records	unique-keys	unique-values		
400	400	200000000	20000	1000000	12.399333	12.194889
800	800	400000000	40000	2000000	20.659444	20.136889
1200	1200	600000000	60000	3000000	26.309222	25.090111
1600	1600	800000000	80000	4000000	35.005000	35.879111
2000	2000	1000000000	100000	5000000	46.440111	41.654889
2400	2400	1200000000	120000	6000000	70.788778	49.962222
2800	2800	1400000000	140000	7000000	78.257444	60.207444
3200	3200	1600000000	160000	8000000	146.859111	69.563667
3600	3600	1800000000	180000	9000000	160.398111	80.417333



从上面的两幅图中我们可以看到，随着 mapper 数量或者 Reducer 数量的增加，hash shuffle 的表现比 sort shuffle 的表现越来越糟糕。

（1） Hash-Based Shuffle 和引入合并机制的 Hash-Based Shuffle



Shuffle write 阶段：在一个上游 stage 结束计算之后，将每一个 task 处理的数据按照 key 进行分区。执行 hash 算法将具有相同的 key 的数据写入到同一个磁盘文件中。每一个磁盘文件对应于 reduce 端的 stage 的一个 task。

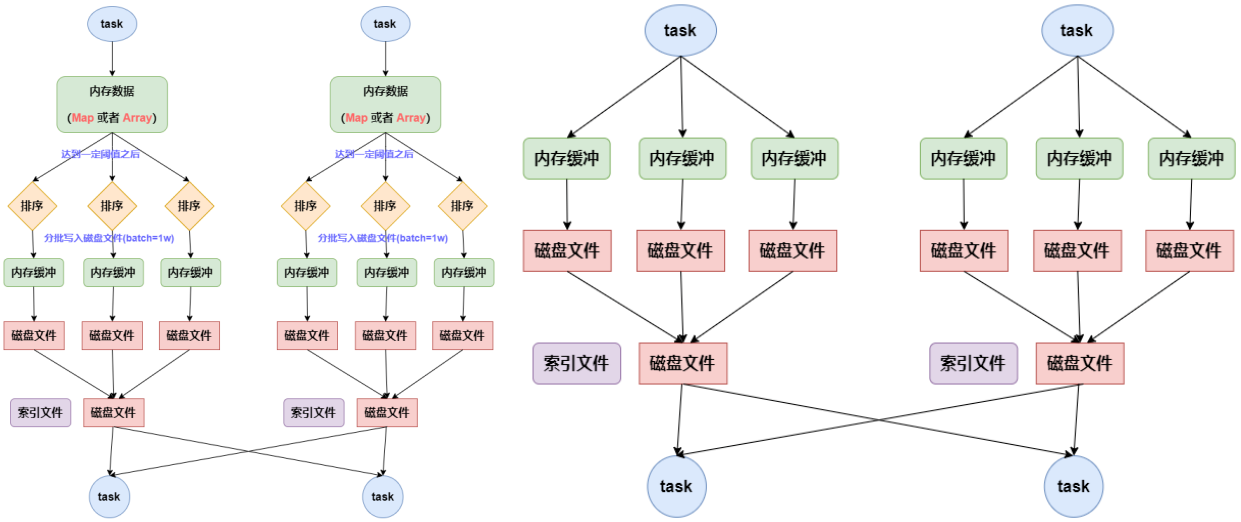
Shuffle read 阶段：由于 write 过程中，每个上游的 task 都给下游每个 task 创建了一个磁盘文件，因此下游 task 直接拉取上游 task 生成的整个磁盘文件。

注：磁盘文件数 = Map Task 数 × Reduce Task 数

开启 consolidate 机制之后，在 write 的过程中，每个 shuffle File Group 为下游 stage 的每个 task 创建一个磁盘文件，而不是每个上游 task 为每个下游 task 创建一个磁盘文件。一个 Executor 上有多少个 CPU Core，就可以并行执行多少个 task，就会创建多少个 shuffle File Group。Executor 的 CPU Core 执行完第一批 task，接着执行下一批 task 时，下一批 task 就会复用之前已有的 shuffle File Group，此时 task 会将数据写入已有的磁盘文件中，而不会写入新的磁盘文件中，这样就可以将多个 task 的磁盘文件进行有效的合并，从而减少磁盘文件的数量，提升 shuffle write 的性能。

注：磁盘文件数 = Executor 数 × Executor 核心数 × Reduce Task 数

（2） Sort-Based Shuffle 和 bypass 机制



排序：在溢写到磁盘文件之前，会先根据 key 对内存中已有的数据进行排序；

溢写：排序过后，会分批将数据写入磁盘文件，数据首先会被缓冲在内存中，当内存缓冲满溢之后再一次写入磁盘文件当中，减少磁盘 IO。

Merge：一个 task 会发生多次磁盘溢写操作，也就会产生多个临时文件。最后会将之前所有的临时文件都进行合并，形成一个磁盘文件和一个索引文件。索引文件标识了下游各个 task 的数据在文件中的 start offset 和 end offset。

一个上游 task 对应一个磁盘文件，也就意味着该 task 为 Reduce 的 task 准备的数据都在这一个文件中。

注：磁盘文件数 = Map Task 数 （索引文件除外）

Sort-Based Shuffle 引入 bypass 机制后，当下游 task 数比较少的情况下，每个 task 都会为每个下游 task 创建一个临时磁盘文件。该过程与普通的 Hash Shuffle 相同，而不会进行排序操作，但是在最后，会将临时磁盘文件合并为一个磁盘文件。相对未经优化的 Hash Shuffle 来说，shuffle read 的性能更好。

四、 设计思路

由以上可知：

Hash Shuffle 磁盘文件数 = Map Task 数 × Reduce Task 数；

Sort Shuffle 磁盘文件数 = Map Task 数；

因此，Hash Shuffle 与 Sort Shuffle 在磁盘文件数上的差异主要由下游分区(task)数决定，当下游 task 数量较多时，shuffle write 阶段将形成大量文件。

同时，Sort Shuffle 为将属于同一 task 的数据合并在一起，需要进行排序，产生额外的计算开销；

实验思路如下：

a) 当下游 task 数增加时，Hash Shuffle 将面临更多的 IO 代价，因此可以通过改变下游分区数，观察 Hash Shuffle 与 Sort Shuffle 的执行性能差异；

b) 当数据规模增加时，Sort Shuffle 将面临更多的计算代价，因此可以通过改变数据规模，观察 Hash Shuffle 与 Sort Shuffle 的执行性能差异；

五、 实验设置

硬件：

CPU：4 核　　内存：16G　　磁盘总容量：100GB

软件：

操作系统：Ubuntu

Hadoop 版本：2.10.1　　Spark 版本：1.6.3

主要参数设置：

数据集：Chinese-Cloze-RC　　规模:2GB

数据集：自动生成的随机数　　规模:1GB—10GB

六、 实验过程

GitHub：

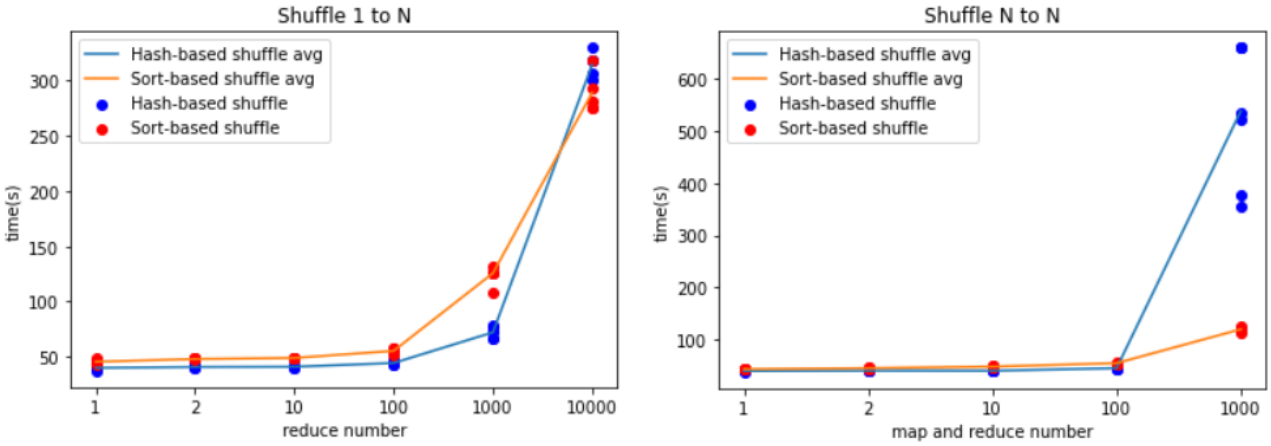
https://github.com/CheenYuee/Distributed-Computing-Systems-Project.git

（1） 针对思路 a)进行验证；

数据来源：Chinese-Cloze-RC 中 pd.train 文件；

测试程序：WordCount。程序中包含 shuffle 类算子——reduceByKey；

实验结果可视化图如下：



图表解释：

1、 左图：横坐标为 reduce 任务数量，纵坐标为执行时间。此图描述的是执行时间随 reduce 任务数量的的增加的变化情况。若横坐标为 N，则 map 任务数量为 1，reduce 的任务数量为 N。

2、 右图：横坐标为 map 任务和 reduce 任务数量，纵坐标为执行时间。此图描述的是执行时间随 map 任务和 reduce 任务数量的增加的变化情况。若横坐标为 N，则 map 任务数量为 N，reduce 任务数量也为 N。

结果解释：

从上可以看出，随着任务数量的增加，执行时间反而增加，说明应用的 IO 代价过高，并且，任务数量越多，IO 代价越高。当 task 数量较少时，由于 Sort Shuffle 排序的性质需要额外的计算开销，Hash Shuffle 的执行速度较快，性能较高。但随着当 task 的数量达到一定值，Hash Shuffle 将逐渐慢于 Sort Shuffle，Hash Shuffle 的

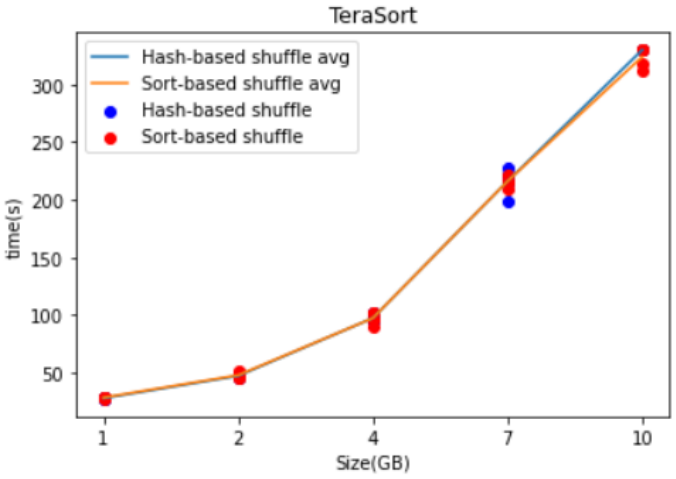
IO 代价将大于 Sort Shuffle。因此，Sort Shuffle 可能将更适用于多任务的应用，其具有更好的 IO 性能。而在一些 task 数量较少的情况下，Hash Shuffle 将快于需要排序开销的 Sort Shuffle。

（2） 针对思路 b)进行验证

数据来源：自动生成的随机数；

测试程序：TeraSort。程序中包含 shuffle 类算子——sortByKey。

实验结果可视化图如下：



图表解释：横坐标为数据规模，纵坐标为执行时间。此图描述的是执行时间随数据规模的增大的变化情况。

结果解释：从上图可以看出，随着数据量的增加，执行时间相应增加，但是 Hash Shuffle 和 Sort Shuffle 的执行效率没有明显差别。因此，在任务数量相同的情况下，Hash Shuffle 和 Sort Shuffle 可能没有显著差别，但是，对于大规模数据的处理，往往需要大量的 Task 任务来并行进行数据的处理。

七、 实验结论

1、当 shuffle 的 task 数量不多时，Hash Shuffle 表现出更好的性能；

2、当 shuffle 的 task 数量较多时，sort Shuffle 表现出更好的性能；

实验结论分析：

Hash Shuffle。对于中小规模数据的处理，Hash Shuffle 具有很好的性能。但当数据量较大，task 数量较多时，Hash Shuffle 将产生大量磁盘文件。当数据量越来越多、task 越来越多时，产生的文件量是不可控的，会对文件系统造成压力。同时，大量小文件的随机读写带来低效的 IO 开销，影响了 Spark 的性能及扩展能力。并且，对于每个文件，都需要在内存中分配一块 buffer，缓存空间的增加将消耗内存空间，甚至出现 out of Memory Error 的情况。因此，针对这种情况，引入了 consolidate 机制。consolidate 机制允许不同的 task 复用同一批磁盘文件，将多个 task 的磁盘文件进行一定程度上的合并，从而大幅度减少磁盘文件的数量，进而提升 shuffle write 的性能。

Sort Shuffle。对于大规模数据处理，Sort Shuffle 通过合并文件的机制减少了最终的磁盘数量，具有较好的性能。但是，Sort Shuffle 排序的特性带来了额外的

计算开销，使得在一些情况下，Sort Shuffle 将会慢于 Hash Shuffle，特别是当下游 task 数较少时，Sort Shuffle 减少的 IO 开销将不足以弥补增加的计算开销，会显得较为低效。因此，针对这种情况，引入了 bypass 机制。当触发 bypass 机制时，shuffle write 过程将不需要进行数据的排序操作，而是采用类似 Hash Shuffle 的方法，提升性能。

总结：

对于中小规模数据处理而言，Hash shuffle 快于 Sorted Shuffle，而 Sorted shuffle 更能胜任大规模数据的处理。但实际上，引入 bypass 机制的 Sort Shuffle 已经吸收了 hash 的优点，并且，Spark 2.0 后加入了更优的 Tungsten-Sort Shuffle 方法。这种方法对排序算法进行了改进，优化了排序的速度，在 2.0 版本后已经并入了 Sorted Shuffle，Spark 引擎会自动识别程序需要的是哪种 shuffle 方法。

附录

错误记录：在实验过程中时常遇到以下错误，特此记录。

错误提示：

ERROR TaskSchedulerImpl: Lost executor 1 on xx.xx.xx.xx: Remote RPC client disassociated. Likely due to containers exceeding thresholds, or network issues. Check driver logs for WARN messages.

解决方法：

spark 节点可用内存不足，在 spark-env.sh 中增大可使用内存；

修改 SPARK_DRIVER_MEMORY 或 SPARK_EXECUTOR_MEMORY 的值；

实验数据截图记录：

WordCount-Hash-Shuffle-1ToN：

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624145236-0019	JavaWordCount-task1	2	14.0 GB	2022/06/24 14:52:36	ubuntu	FINISHED	41 s
app-20220624145149-0018	JavaWordCount-task1	2	14.0 GB	2022/06/24 14:51:49	ubuntu	FINISHED	40 s
app-20220624145102-0017	JavaWordCount-task1	2	14.0 GB	2022/06/24 14:51:02	ubuntu	FINISHED	41 s
app-20220624145015-0016	JavaWordCount-task1	2	14.0 GB	2022/06/24 14:50:15	ubuntu	FINISHED	40 s
app-20220624144935-0015	JavaWordCount-task1	2	14.0 GB	2022/06/24 14:49:35	ubuntu	FINISHED	37 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624145919-0024	JavaWordCount-task2	2	14.0 GB	2022/06/24 14:59:19	ubuntu	FINISHED	40 s
app-20220624145831-0023	JavaWordCount-task2	2	14.0 GB	2022/06/24 14:58:31	ubuntu	FINISHED	41 s
app-20220624145744-0022	JavaWordCount-task2	2	14.0 GB	2022/06/24 14:57:44	ubuntu	FINISHED	41 s
app-20220624145656-0021	JavaWordCount-task2	2	14.0 GB	2022/06/24 14:56:56	ubuntu	FINISHED	41 s
app-20220624145609-0020	JavaWordCount-task2	2	14.0 GB	2022/06/24 14:56:09	ubuntu	FINISHED	40 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624150525-0029	JavaWordCount-task10	2	14.0 GB	2022/06/24 15:05:25	ubuntu	FINISHED	42 s
app-20220624150438-0028	JavaWordCount-task10	2	14.0 GB	2022/06/24 15:04:38	ubuntu	FINISHED	41 s
app-20220624150350-0027	JavaWordCount-task10	2	14.0 GB	2022/06/24 15:03:50	ubuntu	FINISHED	40 s
app-20220624150303-0026	JavaWordCount-task10	2	14.0 GB	2022/06/24 15:03:03	ubuntu	FINISHED	41 s
app-20220624150216-0025	JavaWordCount-task10	2	14.0 GB	2022/06/24 15:02:16	ubuntu	FINISHED	40 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624151140-0034	JavaWordCount-task100	2	14.0 GB	2022/06/24 15:11:40	ubuntu	FINISHED	44 s
app-20220624151046-0033	JavaWordCount-task100	2	14.0 GB	2022/06/24 15:10:46	ubuntu	FINISHED	47 s
app-20220624150956-0032	JavaWordCount-task100	2	14.0 GB	2022/06/24 15:09:56	ubuntu	FINISHED	43 s
app-20220624150906-0031	JavaWordCount-task100	2	14.0 GB	2022/06/24 15:09:06	ubuntu	FINISHED	43 s
app-20220624150816-0030	JavaWordCount-task100	2	14.0 GB	2022/06/24 15:08:16	ubuntu	FINISHED	44 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624151942-0039	JavaWordCount-task1000	2	14.0 GB	2022/06/24 15:19:42	ubuntu	FINISHED	1.1 min
app-20220624151831-0038	JavaWordCount-task1000	2	14.0 GB	2022/06/24 15:18:31	ubuntu	FINISHED	1.1 min
app-20220624151710-0037	JavaWordCount-task1000	2	14.0 GB	2022/06/24 15:17:10	ubuntu	FINISHED	1.3 min
app-20220624151553-0036	JavaWordCount-task1000	2	14.0 GB	2022/06/24 15:15:53	ubuntu	FINISHED	1.2 min
app-20220624151432-0035	JavaWordCount-task1000	2	14.0 GB	2022/06/24 15:14:32	ubuntu	FINISHED	1.3 min

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624154351-0044	JavaWordCount-task10000	2	14.0 GB	2022/06/24 15:43:51	ubuntu	FINISHED	5.0 min
app-20220624153841-0043	JavaWordCount-task10000	2	14.0 GB	2022/06/24 15:38:41	ubuntu	FINISHED	5.1 min
app-20220624153317-0042	JavaWordCount-task10000	2	14.0 GB	2022/06/24 15:33:17	ubuntu	FINISHED	5.3 min
app-20220624152752-0041	JavaWordCount-task10000	2	14.0 GB	2022/06/24 15:27:52	ubuntu	FINISHED	5.3 min
app-20220624152219-0040	JavaWordCount-task10000	2	14.0 GB	2022/06/24 15:22:19	ubuntu	FINISHED	5.5 min

WordCount-Sort-Shuffle-1ToN：

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220622004253-0033	JavaWordCount-task1	2	14.0 GB	2022/06/22 00:42:53	ubuntu	FINISHED	48 s
app-20220622004201-0032	JavaWordCount-task1	2	14.0 GB	2022/06/22 00:42:01	ubuntu	FINISHED	46 s
app-20220622004110-0031	JavaWordCount-task1	2	14.0 GB	2022/06/22 00:41:10	ubuntu	FINISHED	45 s
app-20220622004018-0030	JavaWordCount-task1	2	14.0 GB	2022/06/22 00:40:18	ubuntu	FINISHED	45 s
app-20220622003928-0029	JavaWordCount-task1	2	14.0 GB	2022/06/22 00:39:28	ubuntu	FINISHED	43 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220622005034-0038	JavaWordCount-task2	2	14.0 GB	2022/06/22 00:50:34	ubuntu	FINISHED	48 s
app-20220622004940-0037	JavaWordCount-task2	2	14.0 GB	2022/06/22 00:49:40	ubuntu	FINISHED	48 s
app-20220622004846-0036	JavaWordCount-task2	2	14.0 GB	2022/06/22 00:48:46	ubuntu	FINISHED	47 s
app-20220622004751-0035	JavaWordCount-task2	2	14.0 GB	2022/06/22 00:47:51	ubuntu	FINISHED	48 s
app-20220622004656-0034	JavaWordCount-task2	2	14.0 GB	2022/06/22 00:46:56	ubuntu	FINISHED	48 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624142507-0004	JavaWordCount-task10	2	14.0 GB	2022/06/24 14:25:07	ubuntu	FINISHED	49 s
app-20220624142413-0003	JavaWordCount-task10	2	14.0 GB	2022/06/24 14:24:13	ubuntu	FINISHED	48 s
app-20220624142316-0002	JavaWordCount-task10	2	14.0 GB	2022/06/24 14:23:16	ubuntu	FINISHED	49 s
app-20220624142221-0001	JavaWordCount-task10	2	14.0 GB	2022/06/24 14:22:21	ubuntu	FINISHED	49 s
app-20220624142125-0000	JavaWordCount-task10	2	14.0 GB	2022/06/24 14:21:25	ubuntu	FINISHED	49 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624143310-0009	JavaWordCount-task100	2	14.0 GB	2022/06/24 14:33:10	ubuntu	FINISHED	52 s
app-20220624143210-0008	JavaWordCount-task100	2	14.0 GB	2022/06/24 14:32:10	ubuntu	FINISHED	54 s
app-20220624143108-0007	JavaWordCount-task100	2	14.0 GB	2022/06/24 14:31:08	ubuntu	FINISHED	55 s
app-20220624143004-0006	JavaWordCount-task100	2	14.0 GB	2022/06/24 14:30:04	ubuntu	FINISHED	57 s
app-20220624142900-0005	JavaWordCount-task100	2	14.0 GB	2022/06/24 14:29:00	ubuntu	FINISHED	58 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624144448-0014	JavaWordCount-task1000	2	14.0 GB	2022/06/24 14:44:48	ubuntu	FINISHED	1.8 min
app-20220624144239-0013	JavaWordCount-task1000	2	14.0 GB	2022/06/24 14:42:39	ubuntu	FINISHED	2.1 min
app-20220624144033-0012	JavaWordCount-task1000	2	14.0 GB	2022/06/24 14:40:33	ubuntu	FINISHED	2.1 min
app-20220624143817-0011	JavaWordCount-task1000	2	14.0 GB	2022/06/24 14:38:17	ubuntu	FINISHED	2.2 min
app-20220624143605-0010	JavaWordCount-task1000	2	14.0 GB	2022/06/24 14:36:05	ubuntu	FINISHED	2.1 min

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624160929-0049	JavaWordCount-task10000	2	14.0 GB	2022/06/24 16:09:29	ubuntu	FINISHED	4.6 min
app-20220624160446-0048	JavaWordCount-task10000	2	14.0 GB	2022/06/24 16:04:46	ubuntu	FINISHED	4.7 min
app-20220624155949-0047	JavaWordCount-task10000	2	14.0 GB	2022/06/24 15:59:49	ubuntu	FINISHED	4.9 min
app-20220624155521-0046	JavaWordCount-task10000	2	14.0 GB	2022/06/24 15:55:21	ubuntu	FINISHED	4.4 min
app-20220624155044-0045	JavaWordCount-task10000	2	14.0 GB	2022/06/24 15:50:44	ubuntu	FINISHED	4.6 min

WordCount-Hash-Shuffle-NtoN：

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624171526-0081	JavaWordCount-task1to1	2	14.0 GB	2022/06/24 17:15:26	ubuntu	FINISHED	41 s
app-20220624171439-0080	JavaWordCount-task1to1	2	14.0 GB	2022/06/24 17:14:39	ubuntu	FINISHED	40 s
app-20220624171351-0079	JavaWordCount-task1to1	2	14.0 GB	2022/06/24 17:13:51	ubuntu	FINISHED	41 s
app-20220624171304-0078	JavaWordCount-task1to1	2	14.0 GB	2022/06/24 17:13:04	ubuntu	FINISHED	40 s
app-20220624171220-0077	JavaWordCount-task1to1	2	14.0 GB	2022/06/24 17:12:20	ubuntu	FINISHED	38 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624172035-0086	JavaWordCount-task2to2	2	14.0 GB	2022/06/24 17:20:35	ubuntu	FINISHED	41 s
app-20220624171949-0085	JavaWordCount-task2to2	2	14.0 GB	2022/06/24 17:19:49	ubuntu	FINISHED	40 s
app-20220624171901-0084	JavaWordCount-task2to2	2	14.0 GB	2022/06/24 17:19:01	ubuntu	FINISHED	41 s
app-20220624171813-0083	JavaWordCount-task2to2	2	14.0 GB	2022/06/24 17:18:13	ubuntu	FINISHED	41 s
app-20220624171726-0082	JavaWordCount-task2to2	2	14.0 GB	2022/06/24 17:17:26	ubuntu	FINISHED	40 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624172509-0091	JavaWordCount-task10to10	2	14.0 GB	2022/06/24 17:25:09	ubuntu	FINISHED	41 s
app-20220624172422-0090	JavaWordCount-task10to10	2	14.0 GB	2022/06/24 17:24:22	ubuntu	FINISHED	40 s
app-20220624172334-0089	JavaWordCount-task10to10	2	14.0 GB	2022/06/24 17:23:34	ubuntu	FINISHED	41 s
app-20220624172247-0088	JavaWordCount-task10to10	2	14.0 GB	2022/06/24 17:22:47	ubuntu	FINISHED	40 s
app-20220624172200-0087	JavaWordCount-task10to10	2	14.0 GB	2022/06/24 17:22:00	ubuntu	FINISHED	40 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624173021-0096	JavaWordCount-task100to100	2	14.0 GB	2022/06/24 17:30:21	ubuntu	FINISHED	46 s
app-20220624172929-0095	JavaWordCount-task100to100	2	14.0 GB	2022/06/24 17:29:29	ubuntu	FINISHED	46 s
app-20220624172837-0094	JavaWordCount-task100to100	2	14.0 GB	2022/06/24 17:28:37	ubuntu	FINISHED	45 s
app-20220624172745-0093	JavaWordCount-task100to100	2	14.0 GB	2022/06/24 17:27:45	ubuntu	FINISHED	45 s
app-20220624172654-0092	JavaWordCount-task100to100	2	14.0 GB	2022/06/24 17:26:54	ubuntu	FINISHED	45 s
app-20220624190349-0104	JavaWordCount-task1000to1000	2	14.0 GB	2022/06/24 19:03:49	ubuntu	FINISHED	8.9 min
app-20220624185216-0103	JavaWordCount-task1000to1000	2	14.0 GB	2022/06/24 18:52:16	ubuntu	FINISHED	11 min
app-20220624184543-0102	JavaWordCount-task1000to1000	2	14.0 GB	2022/06/24 18:45:43	ubuntu	FINISHED	6.3 min
app-20220624181025-0101	JavaWordCount-task1000to1000	2	14.0 GB	2022/06/24 18:10:25	ubuntu	FINISHED	35 min
app-20220624175851-0100	JavaWordCount-task1000to1000	2	14.0 GB	2022/06/24 17:58:51	ubuntu	FINISHED	11 min
app-20220624174954-0099	JavaWordCount-task1000to1000	2	14.0 GB	2022/06/24 17:49:54	ubuntu	FINISHED	8.7 min
app-20220624173816-0098	JavaWordCount-task1000to1000	2	14.0 GB	2022/06/24 17:38:16	ubuntu	FINISHED	11 min
app-20220624173209-0097	JavaWordCount-task1000to1000	2	14.0 GB	2022/06/24 17:32:09	ubuntu	FINISHED	5.9 min

WordCount-Sort-Shuffle-NtoN：

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624163940-0056	JavaWordCount-task1to1	2	14.0 GB	2022/06/24 16:39:40	ubuntu	FINISHED	43 s
app-20220624163853-0055	JavaWordCount-task1to1	2	14.0 GB	2022/06/24 16:38:53	ubuntu	FINISHED	44 s
app-20220624163805-0054	JavaWordCount-task1to1	2	14.0 GB	2022/06/24 16:38:05	ubuntu	FINISHED	44 s
app-20220624163718-0053	JavaWordCount-task1to1	2	14.0 GB	2022/06/24 16:37:18	ubuntu	FINISHED	44 s
app-20220624163631-0052	JavaWordCount-task1to1	2	14.0 GB	2022/06/24 16:36:31	ubuntu	FINISHED	43 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624164625-0061	JavaWordCount-task2to2	2	14.0 GB	2022/06/24 16:46:25	ubuntu	FINISHED	46 s
app-20220624164537-0060	JavaWordCount-task2to2	2	14.0 GB	2022/06/24 16:45:37	ubuntu	FINISHED	45 s
app-20220624164448-0059	JavaWordCount-task2to2	2	14.0 GB	2022/06/24 16:44:48	ubuntu	FINISHED	45 s
app-20220624164356-0058	JavaWordCount-task2to2	2	14.0 GB	2022/06/24 16:43:56	ubuntu	FINISHED	47 s
app-20220624164308-0057	JavaWordCount-task2to2	2	14.0 GB	2022/06/24 16:43:08	ubuntu	FINISHED	44 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624165202-0066	JavaWordCount-task10to10	2	14.0 GB	2022/06/24 16:52:02	ubuntu	FINISHED	48 s
app-20220624165106-0065	JavaWordCount-task10to10	2	14.0 GB	2022/06/24 16:51:06	ubuntu	FINISHED	49 s
app-20220624165011-0064	JavaWordCount-task10to10	2	14.0 GB	2022/06/24 16:50:11	ubuntu	FINISHED	48 s
app-20220624164916-0063	JavaWordCount-task10to10	2	14.0 GB	2022/06/24 16:49:16	ubuntu	FINISHED	49 s
app-20220624164821-0062	JavaWordCount-task10to10	2	14.0 GB	2022/06/24 16:48:21	ubuntu	FINISHED	48 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624165758-0071	JavaWordCount-task100to100	2	14.0 GB	2022/06/24 16:57:58	ubuntu	FINISHED	54 s
app-20220624165657-0070	JavaWordCount-task100to100	2	14.0 GB	2022/06/24 16:56:57	ubuntu	FINISHED	54 s
app-20220624165553-0069	JavaWordCount-task100to100	2	14.0 GB	2022/06/24 16:55:53	ubuntu	FINISHED	58 s
app-20220624165451-0068	JavaWordCount-task100to100	2	14.0 GB	2022/06/24 16:54:51	ubuntu	FINISHED	55 s
app-20220624165350-0067	JavaWordCount-task100to100	2	14.0 GB	2022/06/24 16:53:50	ubuntu	FINISHED	54 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624170824-0076	JavaWordCount-task1000to1000	2	14.0 GB	2022/06/24 17:08:24	ubuntu	FINISHED	2.0 min
app-20220624170623-0075	JavaWordCount-task1000to1000	2	14.0 GB	2022/06/24 17:06:23	ubuntu	FINISHED	2.0 min
app-20220624170424-0074	JavaWordCount-task1000to1000	2	14.0 GB	2022/06/24 17:04:24	ubuntu	FINISHED	1.9 min
app-20220624170218-0073	JavaWordCount-task1000to1000	2	14.0 GB	2022/06/24 17:02:18	ubuntu	FINISHED	2.0 min
app-20220624170010-0072	JavaWordCount-task1000to1000	2	14.0 GB	2022/06/24 17:00:10	ubuntu	FINISHED	2.1 min

TeraSort-Hash-Shuffle：

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624194240-0111	TeraSort	2	14.0 GB	2022/06/24 19:42:40	ubuntu	FINISHED	28 s
app-20220624194208-0110	TeraSort	2	14.0 GB	2022/06/24 19:42:08	ubuntu	FINISHED	28 s
app-20220624194136-0109	TeraSort	2	14.0 GB	2022/06/24 19:41:36	ubuntu	FINISHED	28 s
app-20220624194105-0108	TeraSort	2	14.0 GB	2022/06/24 19:41:05	ubuntu	FINISHED	27 s
app-20220624194035-0107	TeraSort	2	14.0 GB	2022/06/24 19:40:35	ubuntu	FINISHED	27 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624212332-0011	TeraSort	2	14.0 GB	2022/06/24 21:23:32	ubuntu	FINISHED	48 s
app-20220624212241-0010	TeraSort	2	14.0 GB	2022/06/24 21:22:41	ubuntu	FINISHED	48 s
app-20220624212150-0009	TeraSort	2	14.0 GB	2022/06/24 21:21:50	ubuntu	FINISHED	47 s
app-20220624212101-0008	TeraSort	2	14.0 GB	2022/06/24 21:21:01	ubuntu	FINISHED	45 s
app-20220624211934-0007	TeraSort	2	14.0 GB	2022/06/24 21:19:34	ubuntu	FINISHED	47 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624220726-0028	TeraSort	2	14.0 GB	2022/06/24 22:07:26	ubuntu	FINISHED	1.6 min
app-20220624220546-0027	TeraSort	2	14.0 GB	2022/06/24 22:05:46	ubuntu	FINISHED	1.6 min
app-20220624220406-0026	TeraSort	2	14.0 GB	2022/06/24 22:04:06	ubuntu	FINISHED	1.6 min
app-20220624220225-0025	TeraSort	2	14.0 GB	2022/06/24 22:02:25	ubuntu	FINISHED	1.6 min
app-20220624220040-0024	TeraSort	2	14.0 GB	2022/06/24 22:00:40	ubuntu	FINISHED	1.7 min

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624222822-0034	TeraSort	2	14.0 GB	2022/06/24 22:28:22	ubuntu	FINISHED	3.8 min
app-20220624222439-0033	TeraSort	2	14.0 GB	2022/06/24 22:24:39	ubuntu	FINISHED	3.6 min
app-20220624222056-0032	TeraSort	2	14.0 GB	2022/06/24 22:20:56	ubuntu	FINISHED	3.6 min
app-20220624221712-0031	TeraSort	2	14.0 GB	2022/06/24 22:17:12	ubuntu	FINISHED	3.7 min
app-20220624221347-0030	TeraSort	2	14.0 GB	2022/06/24 22:13:47	ubuntu	FINISHED	3.3 min

TeraSort-Sort-Shuffle：

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624195246-0121	TeraSort	2	14.0 GB	2022/06/24 19:52:46	ubuntu	FINISHED	28 s
app-20220624195213-0120	TeraSort	2	14.0 GB	2022/06/24 19:52:13	ubuntu	FINISHED	29 s
app-20220624195142-0119	TeraSort	2	14.0 GB	2022/06/24 19:51:42	ubuntu	FINISHED	28 s
app-20220624195109-0118	TeraSort	2	14.0 GB	2022/06/24 19:51:09	ubuntu	FINISHED	29 s
app-20220624195038-0117	TeraSort	2	14.0 GB	2022/06/24 19:50:38	ubuntu	FINISHED	27 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624213159-0017	TeraSort	2	14.0 GB	2022/06/24 21:31:59	ubuntu	FINISHED	48 s
app-20220624213108-0016	TeraSort	2	14.0 GB	2022/06/24 21:31:08	ubuntu	FINISHED	47 s
app-20220624213013-0015	TeraSort	2	14.0 GB	2022/06/24 21:30:13	ubuntu	FINISHED	51 s
app-20220624212923-0014	TeraSort	2	14.0 GB	2022/06/24 21:29:23	ubuntu	FINISHED	46 s
app-20220624212833-0013	TeraSort	2	14.0 GB	2022/06/24 21:28:33	ubuntu	FINISHED	46 s

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624215447-0023	TeraSort	2	14.0 GB	2022/06/24 21:54:47	ubuntu	FINISHED	1.6 min
app-20220624215307-0022	TeraSort	2	14.0 GB	2022/06/24 21:53:07	ubuntu	FINISHED	1.6 min
app-20220624215118-0021	TeraSort	2	14.0 GB	2022/06/24 21:51:18	ubuntu	FINISHED	1.7 min
app-20220624214932-0020	TeraSort	2	14.0 GB	2022/06/24 21:49:32	ubuntu	FINISHED	1.7 min
app-20220624214759-0019	TeraSort	2	14.0 GB	2022/06/24 21:47:59	ubuntu	FINISHED	1.5 min

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624224731-0039	TeraSort	2	14.0 GB	2022/06/24 22:47:31	ubuntu	FINISHED	3.6 min
app-20220624224353-0038	TeraSort	2	14.0 GB	2022/06/24 22:43:53	ubuntu	FINISHED	3.6 min
app-20220624224017-0037	TeraSort	2	14.0 GB	2022/06/24 22:40:17	ubuntu	FINISHED	3.5 min
app-20220624223640-0036	TeraSort	2	14.0 GB	2022/06/24 22:36:40	ubuntu	FINISHED	3.6 min
app-20220624223253-0035	TeraSort	2	14.0 GB	2022/06/24 22:32:53	ubuntu	FINISHED	3.7 min

Completed Applications

Application ID	Name	Cores	Memory per Node	Submitted Time	User	State	Duration
app-20220624201955-0127	TeraSort	2	14.0 GB	2022/06/24 20:19:55	ubuntu	FINISHED	5.2 min
app-20220624201423-0126	TeraSort	2	14.0 GB	2022/06/24 20:14:23	ubuntu	FINISHED	5.5 min
app-20220624200851-0125	TeraSort	2	14.0 GB	2022/06/24 20:08:51	ubuntu	FINISHED	5.5 min
app-20220624200319-0124	TeraSort	2	14.0 GB	2022/06/24 20:03:19	ubuntu	FINISHED	5.5 min
app-20220624195800-0123	TeraSort	2	14.0 GB	2022/06/24 19:58:00	ubuntu	FINISHED	5.3 min
app-20220624195448-0122	TeraGen (10GB)	2	14.0 GB	2022/06/24 19:54:48	ubuntu	FINISHED	1.8 min