Efficient Shuffle Management for DAG Computing Frameworks Based on the FRQ Model

***Abstract*—In industrial large-scale data-parallel analytics, shuffle, namely the cross-network read and the aggregation of partitioned data between tasks with data dependencies, usually bring in large overhead. To reduce shuffle overhead, we present *SCache*, an open-source plug-in system that particularly focuses on shuffle optimization. SCache adopts heuristic pre-scheduling combining with shuffle size prediction to pre-fetch shuffle data and balance load on each node. Meanwhile, SCache takes full advantage of the system memory to accelerate the shuffle process. We also propose a new performance model called *Framework Resources Quantification* (FRQ) model to analyze DAG frameworks and evaluate the SCache shuffle optimization. The FRQ model quantifies the utilization of resources and predicts the execution time of each phase of DAG jobs. We have implemented SCache on both Spark and Hadoop MapReduce. The performance of SCache has been evaluated with both simulations and testbed experiments on a 50-node Amazon EC2 cluster. Those evaluations have demonstrated that, by incorporating SCache, the shuffle overhead of Spark can be reduced by nearly 89%, and the overall completion time of TPC-DS queries improves 40% on average. On Apache Hadoop MapReduce, SCache optimizes end-to-end Terasort completion time by 15%.**

***Index Terms*—Distributed DAG frameworks, Shuffle, Optimization, Performance model**

I. Introduction

WE are in an era of data explosion — 2.5 quintillion bytes of data are created every day according to IBM's report[1]. In *industry 4.0*, an increasing number of IoT sensors are embedded in the industrial production line [1]. During the manufacturing process, the information about the assembly lines, stations, and machines is continuously generated and collected. Using distributed computing frameworks to analyze industrial big data is an inevitable trend [2]. Industrial big data is more structured, correlated, and ready for analytics than traditional big data, because industrial big data is generated by automated equipment [3], [4].

According to a cross-industry study [5], both industrial and traditional big data share a characteristic during analytical processing — a small fraction of the daily workload uses well over 90% of the cluster's resources, and these workloads often contain a huge shuffle size. According to another MapReduce trace analysis from Facebook, the shuffle phase accounts for 33% of the job completion time on average, and up to 70% in shuffle-heavy jobs [6]. The shuffle phase is crucial and heavily affecting the end-to-end application performance. Most of the popular frameworks define jobs as directed acyclic graphs (DAGs), such as map-reduce pipeline in Hadoop MapReduce[2], *RDDs* in Spark[3], vertices in Dryad [7], etc. Shuffle phase is always essential as communication between successive computation stages.

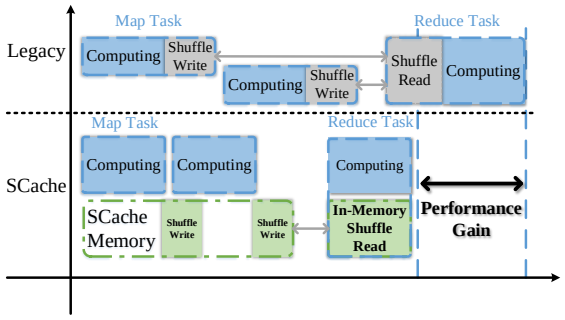
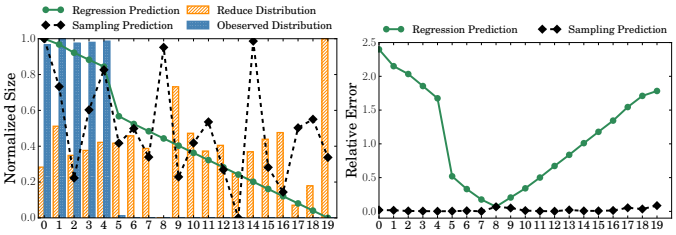


Fig. 1: Workflow Comparison between Legacy DAG Computing Frameworks and Frameworks with SCache

Although continuous efforts of performance optimization have been made among a variety of computing frameworks [8]–[13], the shuffle phase is often poorly optimized in practice. In particular, we observe that one major deficiency lies in the coupled scheduling among different system resources. As Figure 1 shows, the *shuffle write* is responsible for writing intermediate results to disk. And the *shuffle read* fetches intermediate results from remote disks through the network. Once scheduled, a fixed bundle of resources (i.e., CPU, memory, disk, and network) named *slot* is assigned to a task, and *slot*s are released only after the task finishes. Such task aggregation together with the coupled scheduling effectively simplifies task management. However, attaching the I/O intensive shuffle phase to the CPU/memory intensive computation phase results in a poor multiplexing between computational and I/O resources. Moreover, since the shuffle read phase starts fetching data only after the corresponding reduce task starts, all the corresponding reduce tasks start fetching shuffle data almost simultaneously. Such synchronized network communication causes a burst demand for network I/O, which in turn greatly enlarges the shuffle read completion time.

To optimize the data shuffling without significantly changing DAG frameworks, we propose S(huffle)Cache, an open source[4] plug-in system for different DAG computing frameworks. Specifically, SCache takes over the whole shuffle phase from the underlying framework. The workflow of a DAG framework with SCache is presented in Figure 1. SCache replaces the disk operations of shuffle write by the memory copy in map tasks and pre-fetch the shuffle data. SCache's effectiveness lies in the following two key ideas. First, SCache decouples the shuffle write and read from both map and reduce tasks. Such decoupling effectively enables more flexible re-

[1]http://www-01.ibm.com/software/data/bigdata/
[2]http://hadoop.apache.com/
[3]https://spark.apache.org/
[4]https://github.com/frankfzw/SCache



(a) Linear Regression and Sampling Prediction of Range Partitioner

(b) Prediction Relative Error of Range Partitioner

Fig. 2: Reduction Distribution Prediction

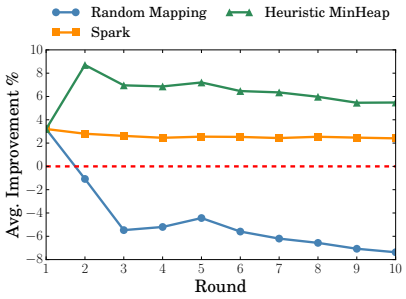


Fig. 3: Stage Completion Time Improvement of OpenCloud Trace

source management and better multiplexing between the computational and I/O resources. Second, SCache pre-schedules the reduce tasks without launching them and pre-fetches the shuffle data. Such pre-scheduling and pre-fetching effectively overlap the network transfer time, desynchronize the network communication, and avoid the extra early allocation of slots.

We evaluate SCache on a 50-node Amazon EC2 cluster on both Spark and Hadoop MapReduce. In a nutshell, SCache can eliminate explicit shuffle time by at most 89% in varied applications. More impressively, SCache reduces 40% of overall completion time of TPC-DS[5], a standardized industry benchmark, on average on Apache Spark.

The rest of the paper is organized as follows: We present the methodologies of optimization which using by SCache in Section II and detail the implementation of SCache in Section III. We introduce the FRQ performance model in Section IV. In Section V, we present comprehensive evaluations about SCache and the FRQ model. We also discuss the related work in Section VI and conclude in Section VII.

II. Shuffle Optimization

This section presents detailed methodologies to achieve shuffle optimization. Firstly, we discuss the reason causes shuffle overhead in the DAG frameworks. In the following subsection, we propose a shuffle data management system to decouple shuffle from execution. And we propose the pre-scheduling and pre-fetching to hide shuffle overhead in multi-round map tasks. Furthermore, two heuristic algorithms (Algorithm 1, 2) are used to improve the accuracy of prediction in the pre-scheduling.

A. Observations

In distributed computing, we use shuffle to do all-to-all data transfer among the nodes. For a clear illustration, we divide a computing job into three phases: map, shuffle, and reduce. In most of the DAG frameworks, including Spark, Hadoop MapReduce, and Tez, shuffle phase is couple with reduce phase which means that shuffle data transfer and reduce computing start simultaneously. On the reduce side, shuffle introduces an explicit overhead during shuffle read. Furthermore, the synchronized shuffle read causes a burst of network traffic and further enlarge the overhead.

[5]http://www.tpc.org/tpcds/

To mitigate the shuffle overhead, we propose an optimization to overlap the shuffle data transfer in multi-round map tasks, and uses memory to cache the shuffle data. To achieve this optimization:

- Shuffle phase should be decoupled from reduce phase to achieve a better scheduling scheme.
- Reduce tasks should be pre-scheduled in map phase to achieve shuffle data pre-fetching.

B. Decouple Shuffle from Execution

To decouple the shuffle from reduce tasks, we propose a shuffle data management system called SCache to take over all shuffle data from the DAG frameworks. SCache provides two APIs named $putBlock$ and $getBlock$ to manage the shuffle data. On the map side, after shuffle data blocks are produced, the map task uses $putBlock$ to transfer the data blocks to SCache. Inside the $putBlock$, SCache uses memory copy to move the shuffle data blocks to SCache's reserved memory and release the slot immediately. After getting the shuffle data, SCache starts to shuffling data immediately. On the reduce side, due to the shuffle data pre-fetching of SCache, the reduce task use $getBlock$ to get the shuffle data from SCache.

SCache cache the shuffle data on the memory instead of storing on disk liked Spark or Hadoop MapReduce. In this aspect, SCache optimizes the time of shuffle read and shuffle write since memory gets better I/O performance.

C. Pre-scheduling and Pre-fetching Shuffle Data

The pre-scheduling and pre-fetching are the most critical aspects of the optimization. SCache starts pre-fetching and hides shuffle overhead into multi-round map tasks. However, the shuffle data cannot be pre-fetched without the awareness of task-node mapping.

To optimize shuffle phase, we propose a co-scheduling scheme with two heuristic algorithms (Algorithm 1, 2). The DAG frameworks follow the co-scheduler to starts reduce tasks.

1) Shuffle Output Prediction: The simplest way of pre-scheduling is mapping tasks to nodes randomly and evenly. But without considering job context, the random mapping scheduling shows poor performance due to data skew and ignoring data locality. For the most DAG applications with random large scale input, the shuffle output can be predicted

accurately by a linear regression model based on the observation that the ratio of map output size and input size are invariant given the same job configuration [14]. However, the linear regression model can fail in some scenarios. For example, some customized partitions may cause large inconsistency between observed map output distribution and the final reduce input distribution. To illustrate the case, we present a particular spark job which using the Spark RangePartitioner in Figure 2a. The observed map outputs are picked randomly. The job uses the Spark RangePartitioner and introduces an extremely high data skew. Due to the extremely high data skew introducing by the partition, the linear regression model cannot fit the result well. That is, for one reduce task, almost all of the input data are produced by a particular map task (e.g., the observed map tasks only produce data for reduce task 0-5 in Figure 2a). The data locality skew results in a missing of other reduce tasks' data in the observed map outputs.

To solve this problem, we propose a new methodology named *weighted reservoir sampling*. SCache uses this method instead of the linear regression to predict output when using a RangePartitioner or a customized non-hash partitioner. For each map task, we use classic reservoir sampling to randomly pick $s \times p$ of samples, where p is the number of reduce tasks and s is a tunable number. After that, the map function is called locally to process the sampled data. Finally, the partitioned outputs are collected with the $InputSize_j$ as the weight of the samples. The $inputSize_j$ is the input size of jth map task. $BlockSize_{ji}$ represents the size of block which is produced by map $task_j$ for reduce $task_i$:

$$BlockSize_{ji} = InputSize_j \times \frac{sample_i}{s \times p} \qquad (1)$$
$$sample_i = \text{number of samples for } reduce_i$$

In Figure 2a, when s is set to 3, the result of sampling prediction is much better than linear regression. Figure 2b further proves that the sampling prediction can provide a much more accurate result than the linear regression.

During both of the predictions, the composition of each reduce partition is calculated as well. We define $prob_i$ as

$$prob_i = \max_{0 \le j \le m} \frac{BlockSize_{kji}}{reduceSize_i} \qquad (2)$$
$$m = \text{number of map tasks}$$

The $reduceSize_i$ represents the size of a reduce task, which is represented by $reduceSize_i = \sum_{j=0}^{m} BlockSize_{ji}$. We use $prob_i$ to achieve a better data locality while performing shuffle pre-scheduling.

2) Heuristic MinHeap Scheduling: To balance load while minimizing the network traffic, we present the *Heuristic MinHeap Scheduling* algorithm (Algorithm 1). To keep the pre-scheduling load balance, we maintain a min-heap in the first *while* loop (i.e., line 6-11). We use the min-heap to simulate the load of each node and applies the longest processing time rule (LPT)[6] to achieve $4/3$-*approximation* optimum. Spark FIFO only can achieve 2-*approximation* optimum. After pre-scheduling, the task-node mapping will be adjusted according to the locality. The $SWAP\_TASKS$ will be triggered when the $host\_id$ of a task does not equal the $assigned\_id$. Based on the $prob$, the normalized probability $norm$ is calculated as a bound of performance degradation. Inside the $SWAP\_TASKS$, tasks will be selected and swapped without exceeding the $upper\_bound$. Since the algorithm only maintains a min-heap and traverses $reduce$ for swapping, the algorithm needs $O(n)$ operations.

To evaluate *Heuristic MinHeap Scheduling* algorithm, we use traces from OpenCloud[7] for the simulation. As shown in Figure 3, we run the simulation under three scheduling schemes: Random Mapping, Spark FIFO, and heuristic MinHeap. After balancing load based on the job context (e.g. shuffle size, data locality), *Heuristic MinHeap Scheduling* has a better improvement (average 5.7%) than Spark (average 2.7%) and Random Mapping.

Algorithm 1 Heuristic MinHeap Scheduling for Single Shuffle

```
1:  procedure SCHEDULE(m, host_ids, p_reduces)
2:      m ← partition number of map tasks
3:      R ← sort p_reduces by size in non-increasing order
4:      M ← min-heap {host_id → ([reduces], size)}
5:      idx ← 0
6:      while idx < lenR do
7:          M[0].size += R[idx].size
8:          M[0].reduces.append(R[idx])
9:          R[idx].assigned_id ← M[0].host_id
10:         Sift down M[0] by size
11:         idx ← idx + 1
12:     max ← maximum size in M
13:     for all reduce in R do                ▷ Heuristic locality swap
14:         if reduce.assigned_id ≠ reduce.host_id then
15:             p ← reduce.prob
16:             norm ← (p − 1/m) / (1 − 1/m) /10
17:             upper_bound ← (1 + norm) × max
18:             SWAP_TASKS(M, reduce, upper_bound)
        return M
19: procedure SWAP_TASKS(M, reduce, upper_bound)
20:     Swap tasks between node host_id and node assigned_id
21:     of reduce without exceeding the upper_bound
22:     of both nodes.
23:     Return if it is impossible.
```

Algorithm 2 Accumulated Heuristic Scheduling for Multi-Shuffles

```
1:  procedure M_SCHEDULE(m, host_id, p_reduces, shuffles)
2:      m ← partition number of map tasks    ▷ shuffles are the
    previous schedule result
3:      for all r in p_reduces do
4:          r.size += shuffles[r.rid].size
5:          new_prob ← shuffles[r.rid].size/r.size
6:          if new_prob ≥ r.prob then
7:              r.prob ← new_prob
8:              r.host_id ← shuffles[r.rid].assigned_host
9:      M ← SCHEDULE(m, host_id, p_reduces)
10:     for all host_id in M do                ▷ Re-shuffle
11:         for all r in M[host_id].reduces do
12:             if host ≠ shuffles[r.rid].assigned_host then
13:                 Re-shuffle data to host
14:                 shuffles[r.rid].assigned_host ← host
        return M
```

[6]http://www.designofapproxalgs.com/

[7]http://ftp.pdl.cmu.edu/pub/datasets/hla/dataset.html

3) Cope with Multiple Shuffle Dependencies: A reduce stage can have more than one shuffle dependency in the current DAG computing frameworks. To cope with multiple shuffle dependencies, we present the *Accumulated Heuristic Scheduling* algorithm. As illustrated in Algorithm 2, the sizes of previous *shuffles* scheduled by *Heuristic MinHeap Scheduling* are counted. When a new shuffle starts, the predicted $size$, $prob$, and $host\_id$ in $p\_reduces$ are accumulated with previous *shuffles*. After scheduling, if the newly $assigned\_id$ of a reduce task did not equal the original one, a re-shuffle will be triggered to transfer data to the new host. This re-shuffle is rare since the previous shuffle data contributes a huge composition (i.e., high $prob$) after the accumulation, which leads to a higher probability of tasks swap in $SWAP\_TASKS$.

To traverse $reduce$ for accumulating previous $shuffle$ and re-shuffling data, the algorithm needs $O(n)$ operations.

III. Implementation

This section presents an overview of the implementation of SCache. We first present the system overview and the detail of sampling in Subsection III-A. The following III-B subsection focuses on two constraints on memory management. In Subsection III-C, we discuss the fault tolerance of the system.

A. System Overview

As shown in Figure 4, SCache consists of three components: a distributed shuffle data management system, a DAG co-scheduler, and a worker daemon. As a plug-in system, SCache needs to rely on a DAG framework. The master node of SCache includes a DAG co-scheduler and a storage manager. The DAG co-scheduler is responsible for pre-scheduling the reduce tasks assignment and then forcing the DAG framework to assign reduce tasks according to this assignment. The storage manager is used for managing the shuffle blocks saved in worker nodes.

When a DAG job is submitted, the DAG information is generated in the framework task scheduler. If the RangePartitioner or a customized non-hash partitioner is used, SCache will insert an extra sampling job before the job starts. The sampling job uses a reservoir sampling algorithm [15] on each partition to gather the sample data. SCache master predicts the reduce partition size based on the data and then call Algorithm 1 or 2 to do the pre-scheduling.

After a map task finishes computing, the shuffle writes in the map task is modified to use SCache API to memory copy all shuffle blocks to the SCache worker reserved memory. The SCache records the block ID and the size and sends them to the master. If the collected shuffle blocks reach a threshold, the SCache co-scheduler will run the scheduling Algorithm 1 or 2 to pre-schedule the reduce tasks. After pre-scheduling, SCache workers start pre-fetching shuffle blocks according to the assignment from the master. To enforce the DAG framework to assign tasks according to the SCache pre-scheduled results, we also insert some lines of codes in the framework scheduler.

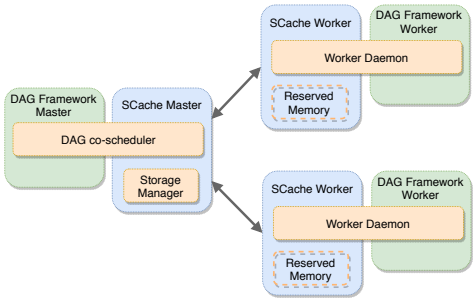


Fig. 4: SCache Architecture

B. Memory Management

SCache uses bellowed memory management to ensure performance. If the reserved memory is running out, SCache will flush some of the blocks into the disk temporarily and then re-fetch them if needed. SCache leverages two constraints to manage the shuffle blocks: *all-or-nothing* and *context-aware-priority*.

1) All-or-Nothing Constraint: As discussed in the observation in Subsection II-A, it is common to have multi-round execution of each stage. If one of the tasks in a round missed a memory cache, the re-fetch overhead will become a bottleneck and then slow down the whole stage. Therefore, SCache master sets blocks of one round as the minimum unit of storage management. We refer to this as the all-or-nothing constraint.

2) Context-Aware-Priority Constraint: SCache leverages application context to select victim storage units when the reserved memory is full. At first, SCache flushes blocks of the incomplete units to disk cluster-widely. If all the units are completed, SCache selects victims based on two factors: (a) For the tasks in the different stages, SCache sets the higher priority to storage units with an earlier submission time; (b) For the tasks in the same stage, SCache sets the higher priority to storage units with a smaller task ID.

C. Discussion of Fault Tolerance

For now, SCache only restarts failed workers without recovering their data. SCache leaves the fault handling to the DAG frameworks. If a failure happens in a SCache worker during shuffle phases, the $getBlock$ API will return a data not found error which causes the DAG framework restarts the current stage. During the re-computing, the DAG frameworks still use SCache to gain shuffle optimization. A possible way to avoid this re-computing is to use replications to recover the data. However, such replications can introduce a significant network overhead. We believe that the current way is more promising because most DAG frameworks have more advanced fast recovery schemes on the application layer, such as the paralleled recovery of Spark.

IV. Framework Resources Quantification Model

In this section, we introduce *Framework Resources Quantification* (FRQ) model to evaluate the performance of DAG frameworks. The FRQ model is able to predict the execution time required by the application under any circumstances,

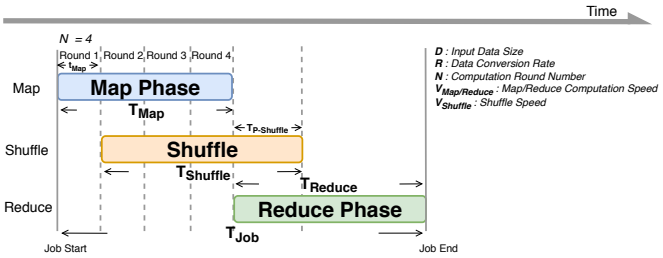


Fig. 5: Framework Resources Quantification Model with Full Parallel MapReduce ($V_{Map} \times R > V_{Shuffle}$)

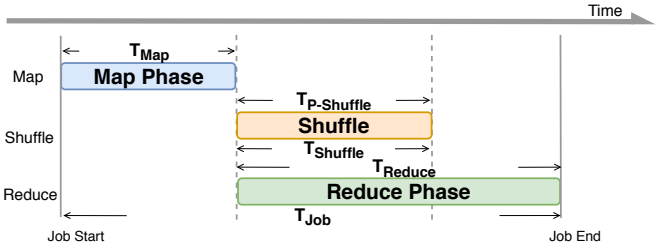


Fig. 6: Framework Resources Quantification Model with Half Parallel MapReduce

including different DAG frameworks, hardware environments, etc. We first introduce the FRQ model in Subsection IV-A. In Subsection IV-B, we use the FRQ model to describe different computation jobs and discuss their performance.

A. The FRQ Model

We propose the FRQ model to better analyze the relationship between the computation phase and the shuffle phase in the DAG computing. The FRQ model focuses on describing the shuffle overhead which significantly effected by the scheduling strategy. After quantifying computing and I/O resources, we use the FRQ model to evaluate different resource scheduling strategies.

As Figure 5 shows, the FRQ model divides the job into three phases: map, shuffle, and reduce. The horizontal axis represents time and the vertical axis represents different phases. Firstly, we introduce five input parameters of the FRQ model:

- Input Data Size (D): The Input data size of the job.
- Data Conversion Rate (R): The conversion rate of the input data to the shuffle data during a computation phase.
- Computation Round Number (N): The number of rounds needed to complete the computation phase. These rounds depend on the current computation resources and the configuration of the framework. Take Hadoop MapReduce as an example. Suppose we have a cluster with 50 CPUs and enough memory, the map phase consists of 200 map tasks, and each map requires 1 CPU. Then we need 4 rounds of computation to complete the map phase.
- Computation Speed (V_i): The computation speed for each computation phase.
- Shuffle Speed ($V_{Shuffle}$): Transmission speed in the shuffle phase.

Besides, the FRQ model also needs to input the scheduling strategies the framework takes. The FRQ model needs to know the start time of each phases. For example, as shown in Figure 6, this strategy starts the shuffle phase and the reduce phase at the same time. And Figure 5 shows the pre-fetching strategy used by SCache. In the pre-fetching strategy, the shuffle phase starts as soon as it can in the map phase.

The FRQ model calculates the execution time of each phase of the job with these five parameters. The total execution time of a job is the sum of the map phase time and reduce phase time:

$$T_{Job} = T_{Map} + T_{Reduce} \tag{3}$$

The formulas of map phase time and shuffle phase time are as follow:

$$T_{Map} = \frac{D}{V_{Map}} \quad T_{Shuffle} = \frac{D}{V_{Shuffle}} \tag{4}$$

The reduce phase time formula is as follows:

$$T_{Reduce} = \frac{D \times R}{V_{Reduce}} + K \times T_{P\_Shuffle} \tag{5}$$

$\frac{D \times R}{V_{Reduce}}$ represents the ideal computation time of a reduce phase, and $(K \times T_{P\_Shuffle})$ represents the computing overhead. $T_{P\_Shuffle}$ represents the overlap time between the shuffle phase and the reduce phase as shown in Figure 5. $T_{Shuffle}$ represents the total time of the shuffle phase. K is an empirical value which represents the overhead caused by shuffle waiting. Because the computation of the reduce phase relies on the data transfer results of the shuffle phase, a portion of the computation needs to wait for the transfer results. This waiting causes the overhead.

According to Equation 3&5, we can optimize the job completion time by reducing $T_{P\_Shuffle}$. By using different resource scheduling strategies, the formulas of $T_{P\_Shuffle}$ are different. In the next subsection, we show the formulas and discuss the performance of different strategies.

B. Different Scheduling Strategies on the FRQ Model

Figure 6 shows a scheduling strategy which is used by Hadoop MapReduce. In this scheduling strategy, shuffle phase and reduce phase start at the same time. In this case, $T_{P\_Shuffle}$ is equal to $T_{Shuffle}$. Due to the increase in $T_{P\_Shuffle}$, the time of reduce phase increases (according to Equation 5). Because the shuffle phase and the computation phase are executed in parallel, the total execution time of a job is the sum of T_{Map} and T_{Reduce} (see Equation 3). The execution time of the shuffle phase is hidden in the reduce phase.

Figure 5 and Figure 7 show the scheduling strategy for Hadoop MapReduce with SCache (Suppose N is 4). SCache also overlaps the shuffle phase and the map phases by starting pre-fetching and pre-scheduling in the map phase. This scheduling strategy avoids the I/O resource being idle in the map phase. According to the design of SCache pre-fetching, we found that using the FRQ model to describe the scheduling strategy of SCache needs to distinguish two situations:

1) $V_{Map} \times R \geq V_{Shuffle}$ (Figure 5)

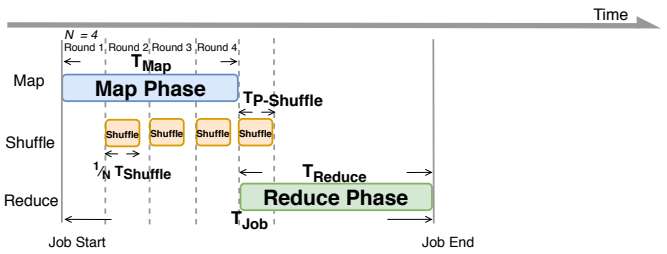
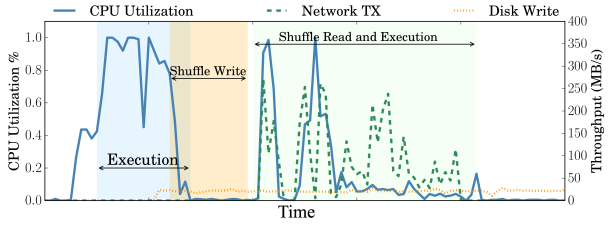
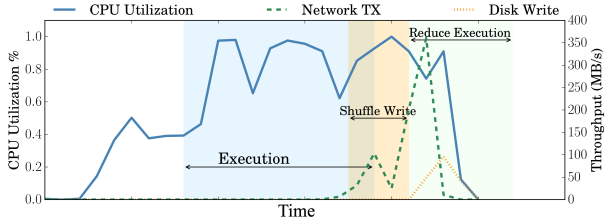


Fig. 7: Framework Resources Quantification Model with Full Parallel MapReduce ($V_{Map} \times R < V_{Shuffle}$)



(a) Spark without SCache



(b) Spark with SCache

Fig. 8: CPU Utilization and I/O Throughput of a Node During a Spark Single Shuffle Application

The meaning of the inequality is that the speed of generating shuffle data ($V_{Map} \times R$) is greater than or equal to the shuffle speed ($V_{Shuffle}$). In this situation, the shuffle phase is uninterrupted. The I/O resource will be fully utilized during the whole shuffle phase. As a result, the formula of $T_{P\_Shuffle}$ is as followed:

$$T_{P\_Shuffle} = T_{Shuffle} - \frac{(N-1) \times T_{Map}}{N} \quad (6)$$

2) $V_{Map} \times R < V_{Shuffle}$ (Figure 7)
When the shuffle speed ($V_{Shuffle}$) is faster, SCache needs to wait for shuffle data to be generated. As Figure 7 shown, the shuffle phase will be interrupted in each round. In this case, the formula of $T_{P\_Shuffle}$ is as followed:

$$T_{P\_Shuffle} = T_{Shuffle} \times \frac{1}{N} \quad (7)$$

Compared to the original Hadoop MapReduce resource scheduling strategy, Hadoop MapReduce with SCache shortens $T_{P\_Shuffle}$ and thus shortens T_{Reduce}. This is how pre-fetching optimizes the total execution time of a job.

V. Evaluation

This section reveals a representative evaluation of SCache performance on both Spark and Hadoop MapReduce. In industrial big data, SQL-based big data technologies are widely used in these systems, such as data mining, predictive analytics, text analytics, and statistical analysis [16]. To prove the performance gain of SCache in industry, we evaluate SCache with TPC-DS[8]. TPC-DS is a standard benchmark which focus on modeling industrial workload. In summary, SCache decreases 89% time of Spark shuffle and improves 40% of the overall completion time of TPC-DS on average. Meanwhile, Hadoop MapReduce with SCache optimizes job completion time by up to 15% and an average of 13%.

A. Setup

We use Spark version 1.6.2 and Hadoop MapReduce version 2.8.5. The shuffle configuration of Spark is set to the default[9]. We run the experiments on a 50-node m4.xlarge cluster on Amazon EC2. Each node has 16GB memory and 4 CPUs. The network bandwidth each node is about 300 Mbps.

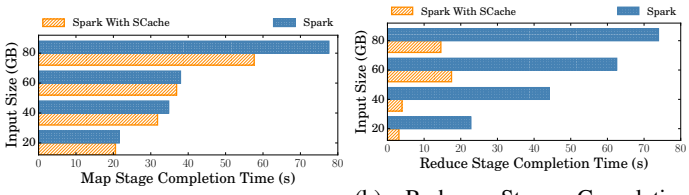
B. Spark with SCache

1) Simple DAG Analysis: We first run Spark's *GroupByTest* on Amazon EC2. This job has 2 rounds of tasks for each node. As shown in Figure 8, the hardware utilization is captured from one node during the job. Note that since the completion time of the job with SCache is about 50% less than Spark without SCache, the duration of Figure 8b is only half of Figure 8a. As shown in 8a, the network transfer of shuffle data introduces an explicit I/O delay during *shuffle read*. And the several bursts of network traffic in *shuffle read* result in network congestion. As shown in Figure 8b, an overlap among CPU, disk, and network can be easily observed. It is because the decoupling of shuffle prevents the computing resource from being blocked by I/O operations. On the one hand, the decoupling of shuffle write helps free the slot earlier, so that it can be re-scheduled to a new map task. On the other hand, with the help of shuffle pre-fetching, the decoupling of shuffle read significantly decreases the CPU idle time at the beginning of a reduce task. At the same time, SCache manages the hardware resources to store and transfer shuffle data without interrupting the computing process. As a result, the utilization and multiplexing of hardware resource are increased, thus improving the performance of Spark.

In the map stage, the disk operations are replaced by the memory copies to decouple the shuffle write. It helps eliminate 40% of shuffle write time (Figure 10a), which leads to a 10% improvement of map stage completion time in Figure 9a. In the reduce stage, most of the shuffle overhead is introduced by network transfer delay. By doing shuffle data pre-fetching, the explicit network transfer is perfectly overlapped in the map stage. As a result, the combination of these optimizations decreases 100% overhead of the shuffle read in a reduce task (Figure 10b). In addition, the heuristic algorithm can achieve a balanced pre-scheduling result, thus providing 80% improvement in reduce stage completion time (Figure 9b).

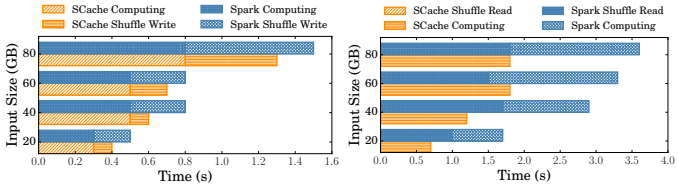
Overall, SCache can help Spark decrease by 89% overhead of the whole shuffle process.

[8]http://www.tpc.org/tpcds/
[9]http://spark.apache.org/docs/1.6.2/configuration.html



(a) Map Stage Completion Time (b) Reduce Stage Completion Time

Fig. 9: Stage Completion Time of Single Shuffle Test



(a) Median Task in Map Stages (b) Median Task in Reduce Stages

Fig. 10: Median Task Completion Time of Single Shuffle Test

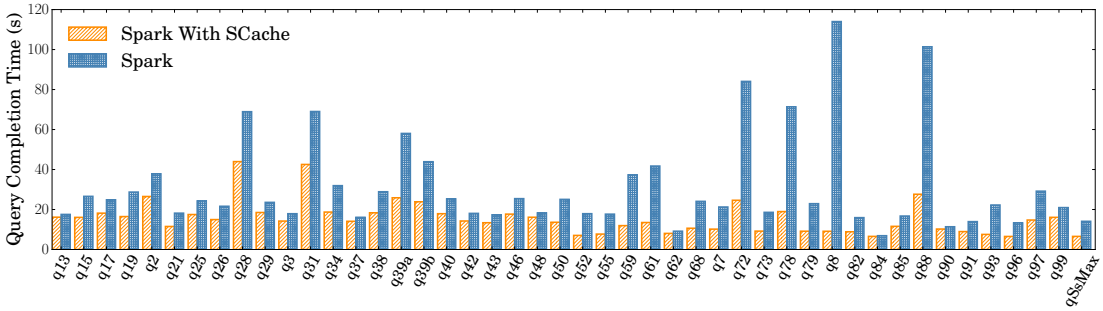
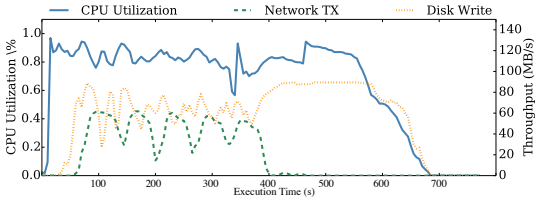
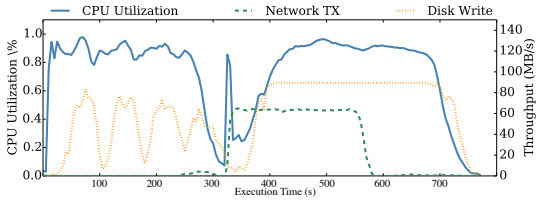


Fig. 11: TPC-DS Benchmark Evaluation



(a) Hadoop MapReduce with SCache



(b) Hadoop MapReduce without SCache

Fig. 12: CPU utilization and I/O throughput of a node during a Hadoop MapReduce Terasort job

2) Industrial Production Workload: TPC-DS[10] contains 99 queries and is considered as a standardized industry benchmark for testing big data systems. As shown in Figure 11, the horizontal axis is query name and the vertical axis is query completion time. Note that we skip some queries due to the compatible issues. Spark with SCache outperforms the original Spark in almost all tested queries. Furthermore, in some shuffle-heavy queries, Spark with SCache outperforms original Spark by an order of magnitude. The overall reduction portion of query time that SCache achieved is 40% on average. Since this evaluation presents the overall job completion time of queries, we believe that our shuffle optimization is promising.

[10]http://www.tpc.org/tpcds/

C. Hadoop MapReduce with SCache

Unlike Spark, Hadoop MapReduce has only one Map and one Reduce phase. Such a simple workflow alleviates the defect of the naive scheduling schemes and reduces the benefits of *pre-scheduling*. However, SCache can still provide considerable optimization on shuffle-heavy workloads. We use Terasort benchmark to evaluate the performance gain of SCache and also prove the cross-framework ability of SCache.

As shown in Figure 12b, Hadoop MapReduce without SCache starts shuffle and reduce simultaneously. The beginning part of reduce phase needs to wait for network transfer because a large amount of shuffle data reaches the network bottleneck. This causes the CPU resources to be idle. On the other hand, in Figure 12a, Hadoop MapReduce with SCache decouples shuffle from reduce and starts pre-fetching in the map phase. SCache utilizes the idle I/O throughput in the map phase. As shown in Figure 13, after decoupling, Hadoop MapReduce with SCache optimizes Terasort overall completion time by up to 15% and an average of 13% with input data sizes from 128GB to 512GB.

D. The FRQ Model Evaluation

To evaluate the FRQ model, we run Terasort with Hadoop MapReduce with two different environments. The parameters D, N, $R, V_{Map}, V_{Shuffle}$,and V_{Reduce} are set according to the application and the environment. We set K to 0.5 and 0.6 since SCache alleviates the defect on the reduce phase Table I shows the results in the in-house environment. While enabling SCache, Terasort satisfies the situation in Figure 5 ($V_{Map} \times R \geq V_{Shuffle}$). While without SCache, $T_{P\_Shuffle}$ is equal to $T_{Shuffle}$(see Equation 4). $ExpT_{Job}$ represents the actual experiment job times and $Error$ represents the error between T_{Job} and $ExpT_{Job}$. Table II shows the results in Amazon EC2.

	D	R	N	V_{Map}	V_{Reduce}	$V_{Shuffle}$	K	T_{Map}	$T_{Shuffle}$	$T_{P\_Shuffle}$	T_{Reduce}	T_{Job}	$ExpT_{Job}$	$Error$
SCache	32 GB	1	4	0.65 GB/s	1 GB/s	0.47 GB/s	0.5	49.23	68.09 s	31.16 s	47.58 s	96.81 s	104 s	6.91%
	48 GB	1	6	0.65 GB/s	1 GB/s	0.47 GB/s	0.5	73.85	102.13 s	40.59 s	68.29 s	142.14 s	151 s	5.87%
	64 GB	1	8	0.65 GB/s	1 GB/s	0.47 GB/s	0.5	98.46	136.17 s	50.02 s	89.01 s	187.47 s	193 s	2.87%
Legacy	32 GB	1	4	0.65 GB/s	1 GB/s	0.47 GB/s	0.6	49.23 s	68.09 s	68.09 s	72.85 s	122.08 s	135 s	9.57%
	48 GB	1	6	0.65 GB/s	1 GB/s	0.47 GB/s	0.6	73.85 s	102.13 s	102.13 s	109.28 s	183.12 s	188 s	2.59%
	64 GB	1	8	0.65 GB/s	1 GB/s	0.47 GB/s	0.6	98.46 s	136.17 s	136.17 s	145.70 s	244.16 s	249 s	1.94%

TABLE I: Hadoop MapReduce on 4 nodes cluster in the FRQ model

	D	R	N	V_{Map}	V_{Reduce}	$V_{Shuffle}$	K	T_{Map}	$T_{Shuffle}$	$T_{P\_Shuffle}$	T_{Reduce}	T_{Job}	$ExpT_{Job}$	$Error$
SCache	128 GB	1	5	1.15 GB/s	1.46 GB/s	1.4 GB/s	0.5	111.30	91.43	18.29	96.81	208.12	232	10.29%
	256 GB	1	5	1.15 GB/s	1.46 GB/s	1.4 GB/s	0.5	222.61	182.86	36.57	193.63	416.24	432	3.65%
	384 GB	1	5	1.15 GB/s	1.46 GB/s	1.4 GB/s	0.5	333.91	274.29	54.86	290.44	624.36	685	8.85%
Legacy	128 GB	1	5	1.15 GB/s	1.46 GB/s	1.4 GB/s	0.6	111.30 s	91.43 s	91.43 s	142.53 s	253.83 s	266 s	4.57%
	256 GB	1	5	1.15 GB/s	1.46 GB/s	1.4 GB/s	0.6	222.61 s	182.86 s	182.86 s	285.06 s	507.67 s	524 s	3.12%
	384 GB	1	5	1.15 GB/s	1.46 GB/s	1.4 GB/s	0.6	333.91 s	274.29 s	274.29 s	427.59 s	761.50 s	776 s	1.87%

TABLE II: Hadoop MapReduce on 50 AWS m4.xlarge nodes cluster in the FRQ model

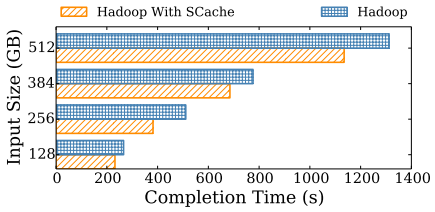


Fig. 13: Hadoop MapReduce Terasort Completion Time

In this environment, Terasort on Hadoop MapReduce satisfies the situation in Figure 7 ($V_{Map} \times R < V_{Shuffle}$), thus the formula of $T_{P\_Shuffle}$ is Equation 7.

As shown in Table I&II, the experimental values are all larger than the calculated values. This is because the application has some extra overhead at runtime, such as network warm-up, the overhead of allocating slots, etc. This overhead will be amplified when the input data is small or the total execution time is short. Overall, the error between T_{Job} and $ExpT_{Job}$ is mainly below 10%, such errors are acceptable.

VI. Related Work

Modeling. Khan et al. [17] proposed the HP model which extends the ARIA model. The HP model adds scaling factors and uses a simple linear regression to estimate the job execution time on larger datasets. Chen et al. [18] proposed the CRESP model which is a cost model that estimates the performance of a job then Farshid Farhat et al. [19] proposed a closed-form queuing model which focuses on stragglers and try to optimize them. However, the above models are not able to accurately describe the overhead caused by the shuffle process under different scheduling strategies.

Industrial big data. More than 1000 Exabytes industrial data is annually generated by smart factories and the data is expected to increase 20-fold in the next ten years [20]. Nowadays, the real challenge of big data is not how to collect it, but how to manage it logically and efficiently [4]. Several distributed computing frameworks become efficient tools in industrial big data analysis [1], [21].

DAG Optimization. Slow-start from Hadoop MapReduce is a classic approach to handle shuffle overhead. Starfish [22] gets sampled data statics for self-tuning system parameters (e.g. slow-start, etc). iShuffle [14] decouples shuffle from reducers and designs a centralized shuffle controller. But it can neither handle multiple shuffles nor schedule multiple rounds of reduce tasks. iHadoop [23] aggressively pre-schedules tasks in multiple successive stages to start fetching shuffle. In data skew mitigation, Skewtune [24] partitions the data dynamically. When a straggler task happens, it re-partitions the data and forces them to transfer to other nodes. LIBRA [25] proposes a new sampling method for DAG frameworks and uses an innovative approach, which supports the split of large keys, to balance the load among the reduce tasks. ToF [26] proposed a DAG optimization system to address the performance and monetary cost optimizations in the cloud. Different from these works, SCache pre-schedules multiple shuffles and uses pre-fetching to gain optimization without breaking load balancing.

VII. Conclusion

In this paper, we present SCache, a plug-in shuffle optimization for DAG computing frameworks. By task pre-scheduling and shuffle data pre-fetching with application context, SCache decouples the shuffle from computing tasks and mitigates the shuffle overhead. We also propose *Framework Resources Quantification* (FRQ) model to evaluate the shuffle process of DAG computing frameworks. The evaluations show that SCache can provide a promising speedup in industrial production workloads.

References

[1] P. Lade, R. Ghosh, and S. Srinivasan, "Manufacturing analytics and industrial internet of things," IEEE Intelligent Systems, vol. 32, no. 3, pp. 74–79, 2017.

[2] J. Lee, H.-A. Kao, and S. Yang, "Service innovation and smart analytics for industry 4.0 and big data environment," Procedia Cirp, vol. 16, pp. 3–8, 2014.

[3] P. Basanta-Val, "An efficient industrial big-data engine," IEEE Transactions on Industrial Informatics, vol. 14, no. 4, pp. 1361–1369, 2018.

[4] Z. Lv, H. Song, P. Basanta-Val, A. Steed, and M. Jo, "Next-generation big data analytics: State of the art, challenges, and future research topics," IEEE Transactions on Industrial Informatics, vol. 13, no. 4, pp. 1891–1899, 2017.

[5] Y. Chen, S. Alspaugh, and R. Katz, "Interactive analytical processing in big data systems: A cross-industry study of mapreduce workloads," Proceedings of the VLDB Endowment, vol. 5, no. 12, pp. 1802–1813, 2012.

[6] M. Chowdhury, M. Zaharia, J. Ma, M. I. Jordan, and I. Stoica, "Managing data transfers in computer clusters with orchestra," in ACM SIGCOMM Computer Communication Review, vol. 41, no. 4. ACM, 2011.

[7] M. Isard, M. Budiu, Y. Yu, A. Birrell, and D. Fetterly, "Dryad: distributed data-parallel programs from sequential building blocks," in ACM SIGOPS operating systems review, vol. 41, no. 3. ACM, 2007.

[8] J. Chen, K. Li, Z. Tang, K. Bilal, S. Yu, C. Weng, and K. Li, "A parallel random forest algorithm for big data in a spark cloud computing environment," IEEE Transactions on Parallel and Distributed Systems, vol. 28, no. 4, pp. 919–933, 2017.

[9] C. Xie, R. Chen, H. Guan, B. Zang, and H. Chen, "Sync or async: Time to fuse for distributed graph-parallel computation," in Proceedings of the 20th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming, ser. PPoPP 2015. New York, NY, USA: ACM, 2015.

[10] H. Li, A. Ghodsi, M. Zaharia, S. Shenker, and I. Stoica, "Tachyon: Reliable, memory speed storage for cluster computing frameworks," in Proceedings of the ACM Symposium on Cloud Computing, 2014.

[11] B. Heintz, A. Chandra, R. K. Sitaraman, and J. Weissman, "End-to-end optimization for geo-distributed mapreduce," IEEE Transactions on Cloud Computing, vol. 4, no. 3, pp. 293–306, 2016.

[12] D. Cheng, J. Rao, Y. Guo, C. Jiang, and X. Zhou, "Improving performance of heterogeneous mapreduce clusters with adaptive task tuning," IEEE Transactions on Parallel and Distributed Systems, vol. 28, no. 3, pp. 774–786, 2017.

[13] M. Wasi-ur Rahman, N. S. Islam, X. Lu, and D. K. D. Panda, "A comprehensive study of mapreduce over lustre for intermediate data placement and shuffle strategies on hpc clusters," IEEE Transactions on Parallel and Distributed Systems, vol. 28, no. 3, pp. 633–646, 2017.

[14] Y. Guo, J. Rao, D. Cheng, and X. Zhou, "ishuffle: Improving hadoop performance with shuffle-on-write," IEEE Transactions on Parallel and Distributed Systems, vol. 28, no. 6, pp. 1649–1662, 2017.

[15] J. S. Vitter, "Random sampling with a reservoir," ACM Transactions on Mathematical Software (TOMS), vol. 11, no. 1, pp. 37–57, 1985.

[16] M. Poess, T. Rabl, and H.-A. Jacobsen, "Analysis of tpc-ds: the first standard benchmark for sql-based big data systems," in Proceedings of the 2017 Symposium on Cloud Computing. ACM, 2017.

[17] M. Khan, Y. Jin, M. Li, Y. Xiang, and C. Jiang, "Hadoop performance modeling for job estimation and resource provisioning," IEEE Transactions on Parallel & Distributed Systems, no. 2, pp. 441–454, 2016.

[18] K. Chen, J. Powers, S. Guo, and F. Tian, "Cresp: Towards optimal resource provisioning for mapreduce computing in public clouds," IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 6, pp. 1403–1412, 2014.

[19] F. Farhat, D. Tootaghaj, Y. He, A. Sivasubramaniam, M. Kandemir, and C. Das, "Stochastic modeling and optimization of stragglers," IEEE Transactions on Cloud Computing, 2016.

[20] S. Yin and O. Kaynak, "Big data for modern industry: challenges and trends [point of view]," Proceedings of the IEEE, vol. 103, no. 2, pp. 143–146, 2015.

[21] X. Li, J. Song, and B. Huang, "A scientific workflow management system architecture and its scheduling based on cloud service platform for manufacturing big data analytics," The International Journal of Advanced Manufacturing Technology, vol. 84, no. 1-4, pp. 119–131, 2016.

[22] H. Herodotou, H. Lim, G. Luo, N. Borisov, L. Dong, F. B. Cetin, and S. Babu, "Starfish: A self-tuning system for big data analytics." in Cidr, vol. 11, no. 2011, 2011.

[23] E. Elnikety, T. Elsayed, and H. E. Ramadan, "ihadoop: asynchronous iterations for mapreduce," in Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on, 2011.

[24] Y. Kwon, M. Balazinska, B. Howe, and J. Rolia, "Skewtune: mitigating skew in mapreduce applications," in Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, 2012.

[25] Q. Chen, J. Yao, and Z. Xiao, "Libra: Lightweight data skew mitigation in mapreduce," IEEE Transactions on parallel and distributed systems, vol. 26, no. 9, 2014.

[26] A. C. Zhou and B. He, "Transformation-based monetary costoptimizations for workflows in the cloud," IEEE Transactions on Cloud Computing, vol. 2, no. 1, pp. 85–98, 2014.