

Reducer Capacity and Communication Cost in MapReduce Algorithms Design

Foto Afrati¹, Shlomi Dolev², Ephraim Korach², Shantanu Sharma², and Jeffrey D. Ullman³

¹National Technical University of Athens, Greece. ²Ben-Gurion University of the Negev, Israel. ³Stanford University, USA.

Goal: Minimizing the communication cost in a MapReduce job

1 Communication Cost

- The total amount of data to transfer from the map phase to the reduce phase.
- Dominates the performance** of a MapReduce algorithm.

2 Reducer Capacity, q

- Reducers do **not** have an **unbounded memory**.
- An **upper bound on the sum of the sizes of the inputs** that are assigned to the reducer.

3 Mapping Schema

An assignment of the set of inputs to some given reducers such that

- A reducer is assigned inputs whose sum of the sizes is less than or equal to the reducer capacity.
- For each output, must assign the corresponding inputs to at least one reducer in common.

4 All-to-All Mapping Schema Problem

- Inputs:** A list of inputs
- Outputs:** Each pair of inputs corresponds to one output
- Example:** Similarity-join

Inputs $w_1 = w_2 = w_3 = 0.20q$, $w_4 = w_5 = 0.19q$, $w_6 = w_7 = 0.18q$

Assignment of inputs

w_1, w_2	w_3, w_4	w_3, w_4	w_5, w_6
w_1, w_2	w_5, w_6	w_3, w_4	w_7
w_1, w_2	w_7	w_5, w_6	w_7

6 reducers
& non-optimum communication cost

Assignment of inputs

w_1, w_2, w_3, w_4, w_7
w_1, w_2, w_5, w_6, w_7
w_3, w_4, w_5, w_6, w_7

3 reducers
& optimum communication cost

5 X-to-Y Mapping Schema Problem

- Inputs:** Two sets X and Y
- Outputs:** Each pair of inputs $\langle x_i, y_i \rangle$, $\forall x_i \in X, \forall y_i \in Y$
- Example:** Skewjoin

Set X : $w_1 = w_2 = w_3 = w_4 = 0.25q$

Set Y : $w'_1 = w'_2 = 0.25q$, $w'_3 = w'_4 = 0.24q$, $w'_5 = w'_6 = 0.23q$

Assignment of inputs

w_1	w'_1, w'_2, w'_3	w_3	w'_1, w'_2, w'_3
w_1	w'_4, w'_5, w'_6	w_3	w'_4, w'_5, w'_6
w_2	w'_1, w'_2, w'_3	w_4	w'_1, w'_2, w'_3
w_2	w'_4, w'_5, w'_6	w_4	w'_4, w'_5, w'_6

8 reducers & non-optimum communication cost

Assignment of inputs

w_1, w_2	w'_1, w'_2	w_3, w_4	w'_1, w'_2
w_1, w_2	w'_3, w'_4	w_3, w_4	w'_3, w'_4
w_1, w_2	w'_5, w'_6	w_3, w_4	w'_5, w'_6

6 reducers
& optimum communication cost

6 Tradeoffs

- The reducer capacity v/s the total number of reducers
- The reducer capacity v/s the parallelism at the reduce phase
- The reducer capacity v/s the communication cost

7 Reference

- F. Afrati, S. Dolev, E. Korach, S. Sharma, and J.D. Ullman. *Assignment of different-sized inputs in MapReduce*. In 2nd Workshop on Algorithms and Systems for MapReduce and Beyond (BeyondMR), pages 1–10, 2015.

