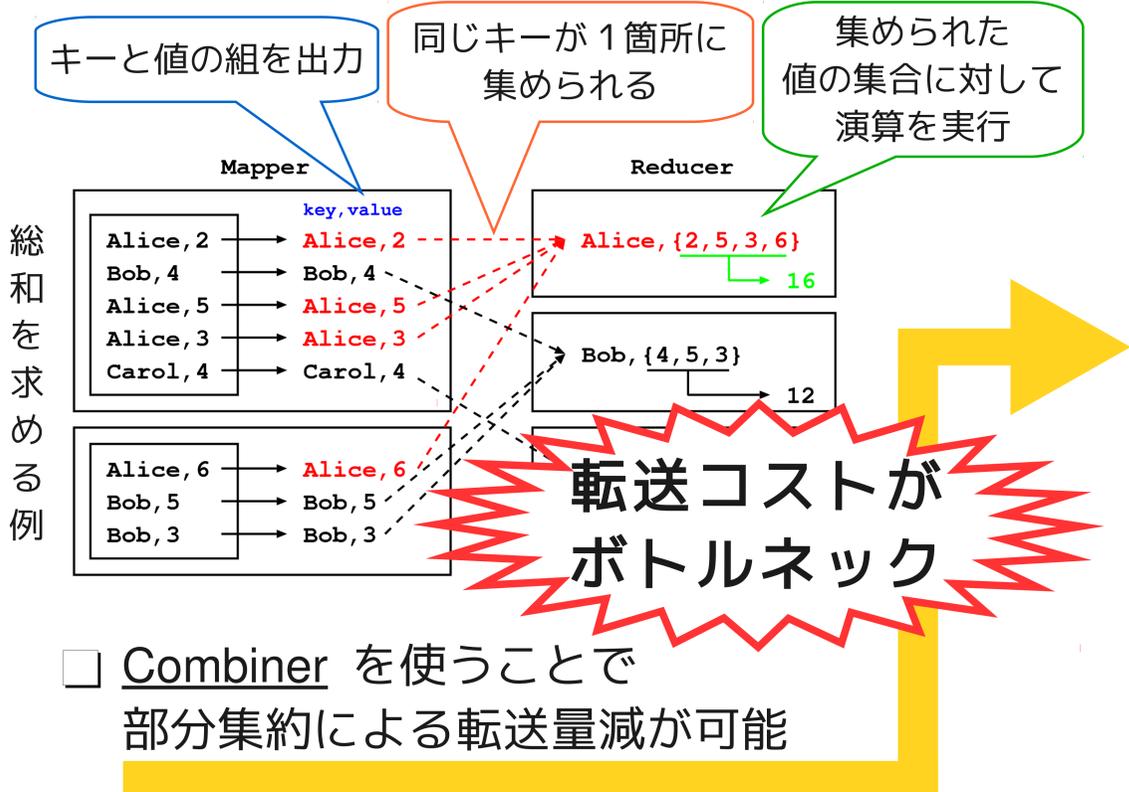


MapReduce フレームワークにおける Combiner の自動生成

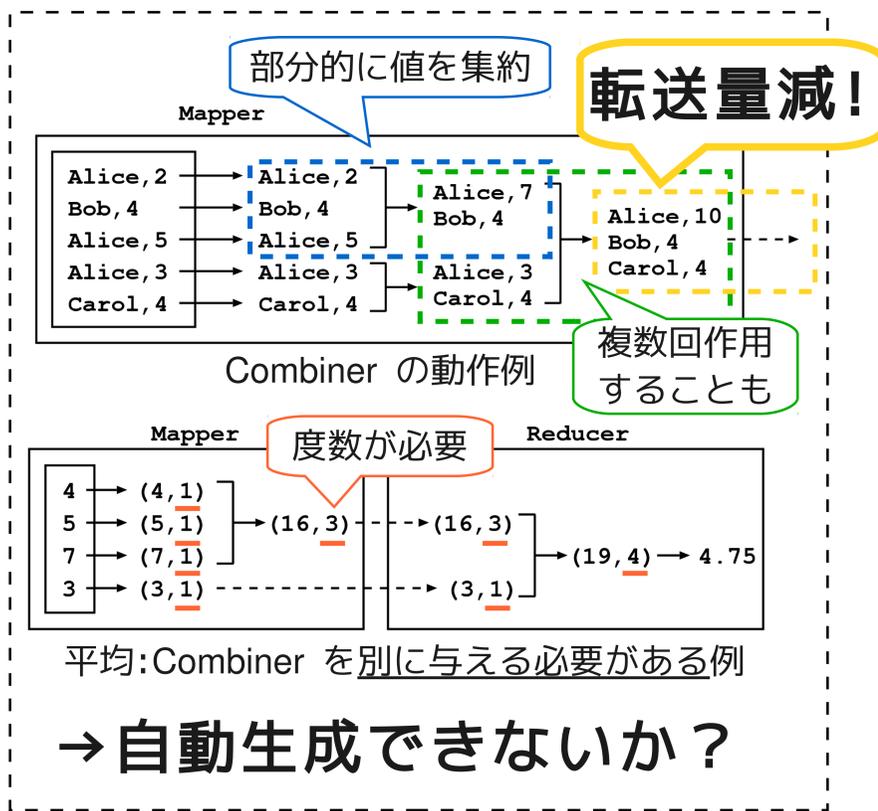
京都大学 樹下 稔、末永 幸平、五十嵐 淳

MapReduce

- 並列計算を簡潔に記述できるフレームワーク
- Map と Reduce の2つのステップで計算を行う



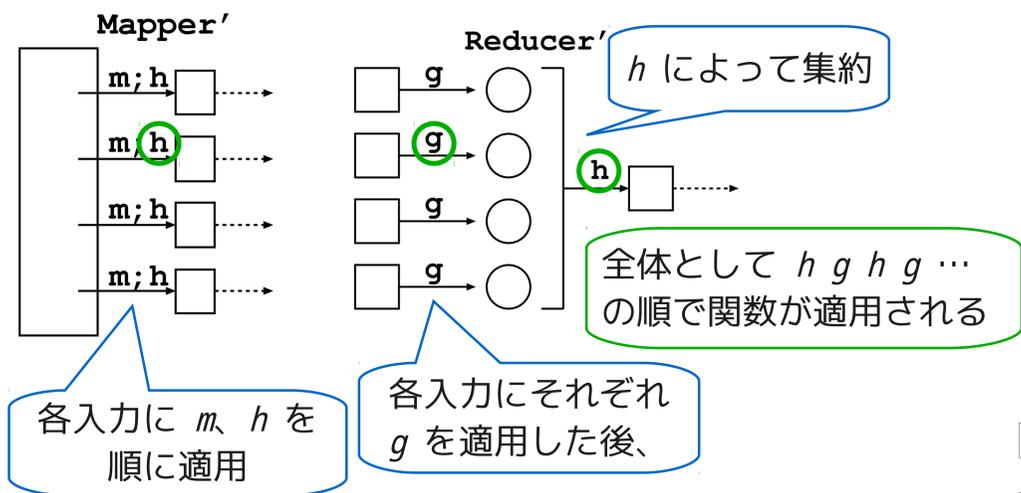
図解: Combiner



生成手法

入力: Mapper m と Reducer h

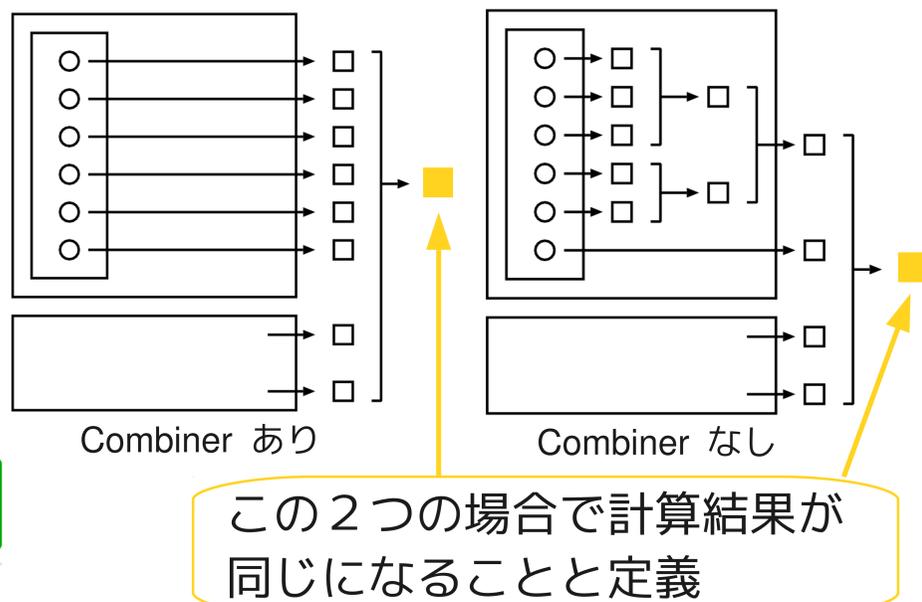
1. h の右逆関数 g を求める
 - $h \circ g \circ h = h$ が成立
2. h と g から Mapper' と Reducer' を生成



3. Combiner と Reducer に Reducer' を使う

正当性の証明

- Combiner の正当性とは?



- 計算過程を木構造としてモデル化
- 木の構造に関する帰納法で証明

実験

- Reducer から Hadoop 用の Java コードを生成するプログラムを作成

環境 Amazon Elastic MapReduce

ノード数 10

入力 10^8 個の 32 ビット整数

問題 Maximum Prefix Sum

	転送量 \propto データ数	実行時間 (秒)
Combiner なし	10^8	282.542
Combiner あり	10^2	153.548

Mapper から Reducer への