Word count example (naive version)

Mapper 1
apple, apple, peach
apple, apple, peach
apple, apple, apple

Mapper 2
peach, peach, apple
peach, peach, apple
peach, peach, banana

Reducer
apple, apple, peach
apple, apple, peach
apple, apple, apple
peach, peach, apple
peach, peach, apple
peach, peach, banana

SUM() is associative and commutative (AC), therefore it can be pushed up.

However, AC is a strong condition, and not many aggregation functions are AC. Question: given an arbitrary aggregation $\alpha()$, how to process it?

Slow: all data are shuffled on a network!!!

Fast: less data are shuffled on a network!!!