

# Comparing Word Count

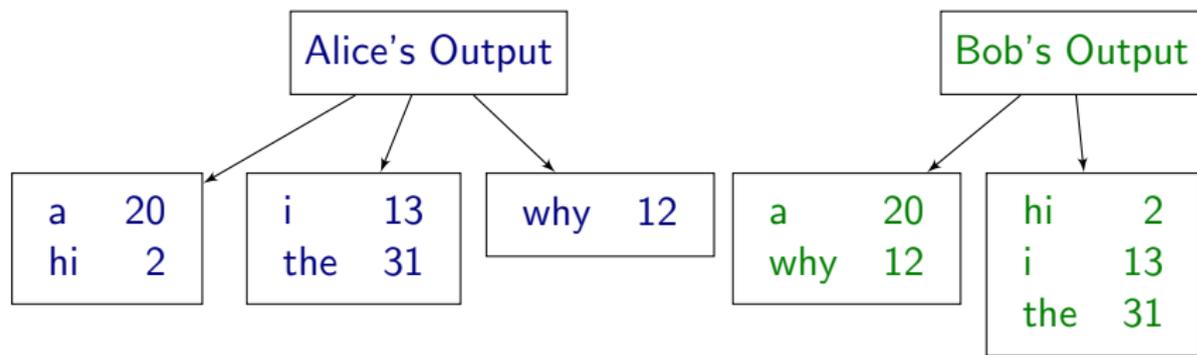
## Problem

Do Alice and Bob agree on their word count output?

## Goals

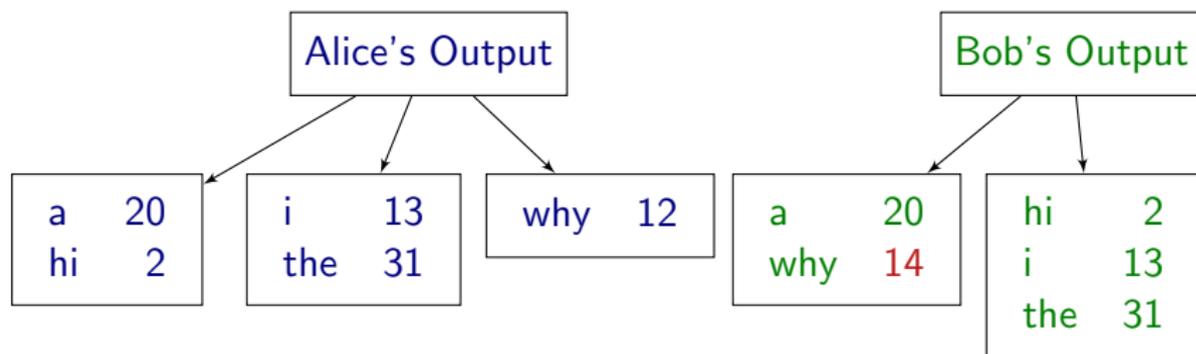
- Identify differences
- Fast

# Input



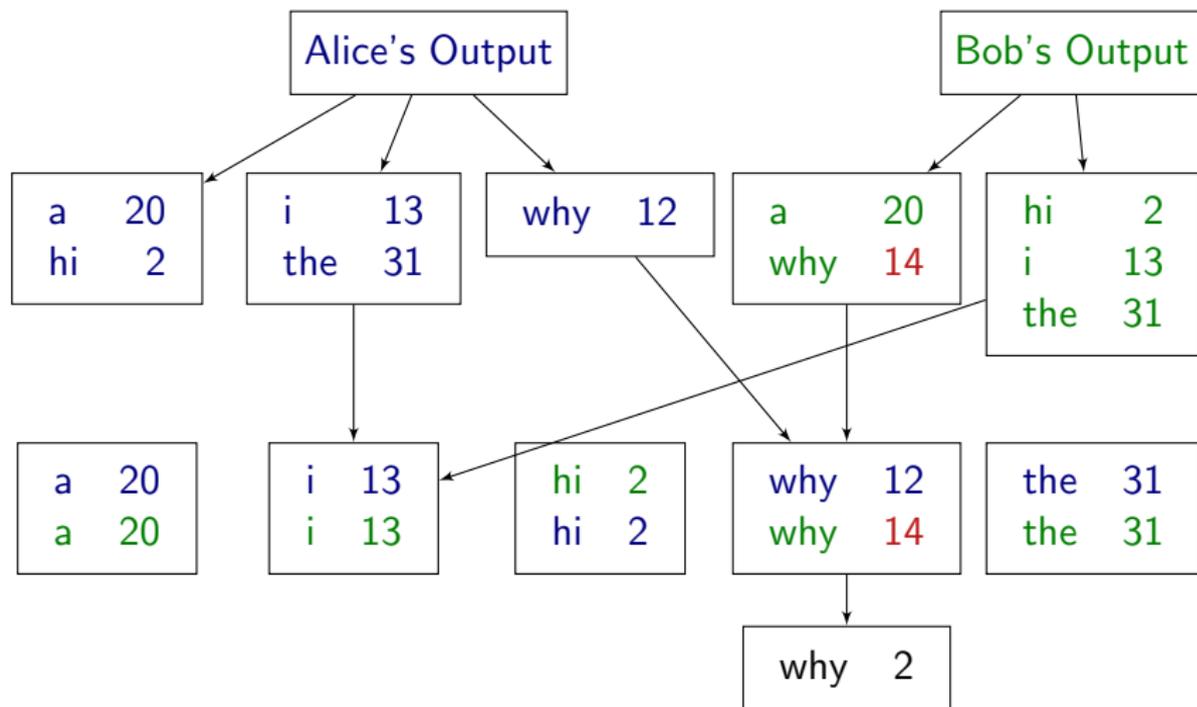
- Alice chose to run 3 reducers, resulting in 3 parts
- Bob chose to run 2 reducers, resulting in 2 parts
- Alice and Bob agree on their outputs
- Files are sorted

# Input

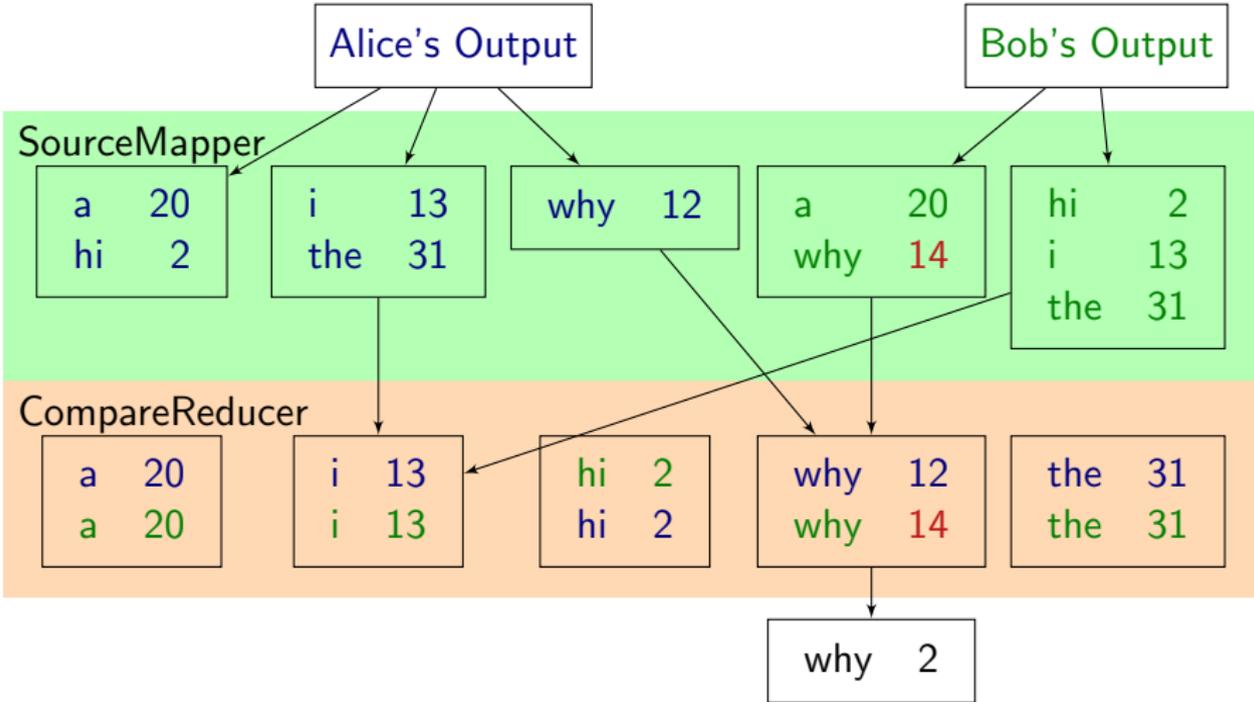


- Now **Bob** counted 2 more of “why”

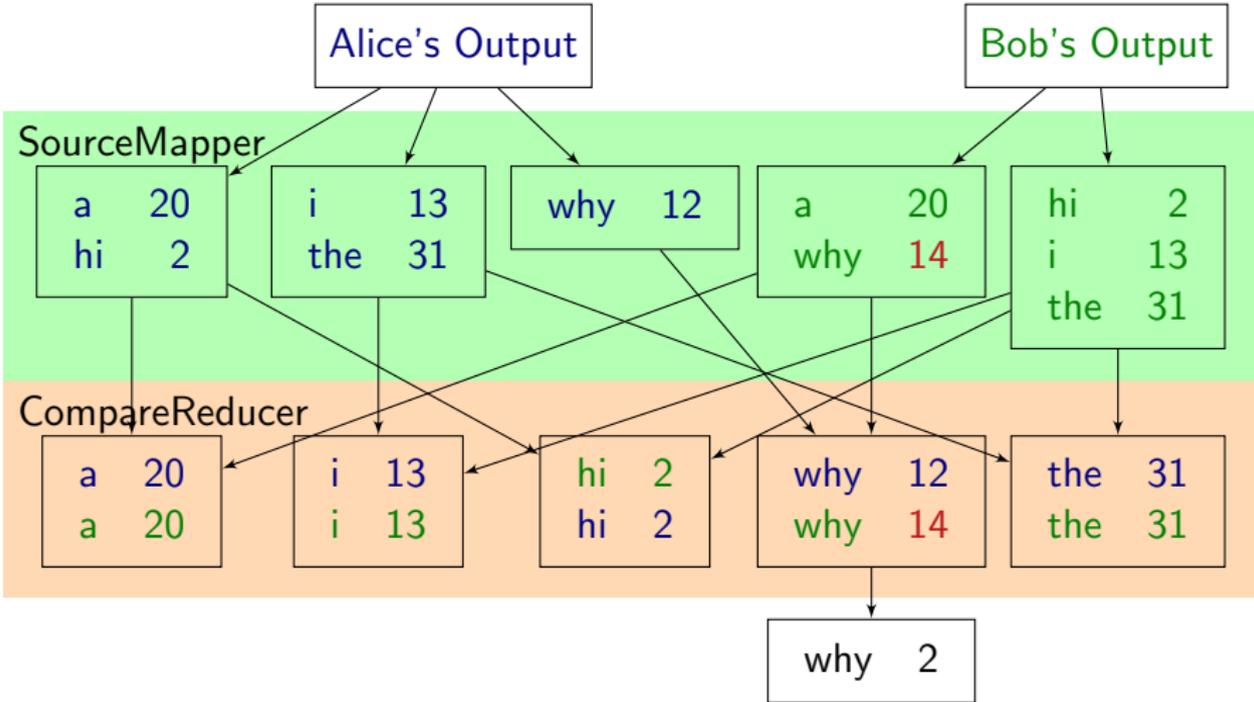
# Output



# MapReduce



# MapReduce



# Reducer Joins

## Implementation

- Input files named `alice00000`, `bob00000`, etc.
- Mapper determines if it is reading Alice or Bob by the file name
- Map input key is *word* and value is *count*.
  - Alice Output key *word* and value (Alice, *count*).
  - Bob Output key *word* and value (Bob, *count*).
- Reducer sees key *word* and values (Alice, *count*) and (Bob, *count*)
  - But not necessarily in order

## Issues

- Copies data to Mapper and again to Reducer.