

# Hadoop MapReduce

Chu Zihao

Deng Ruofan

Wang Fu

**Idea**

**01**

**Demo**

**02**

**Results**

**03**

**Improvements**

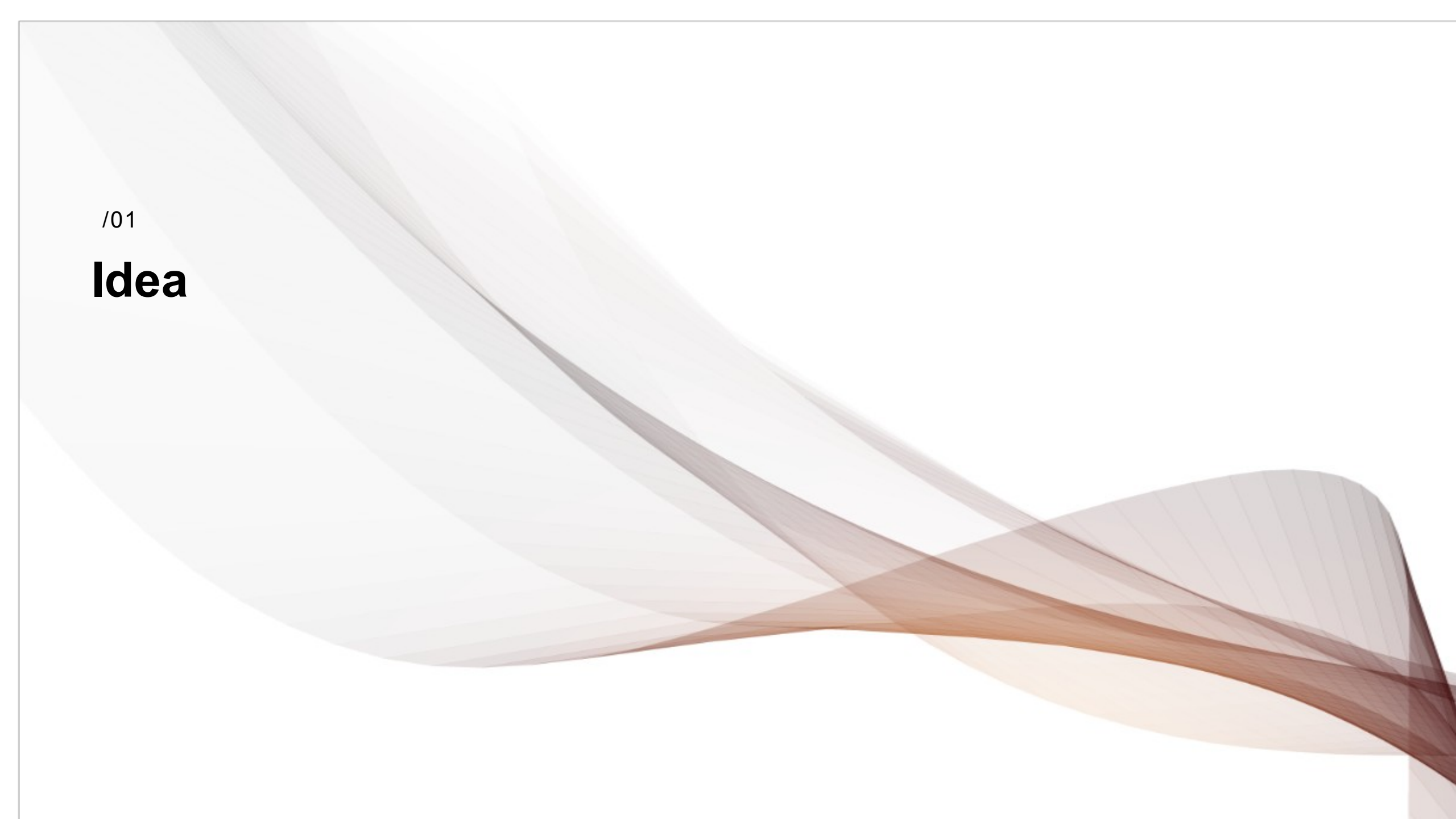
**04**

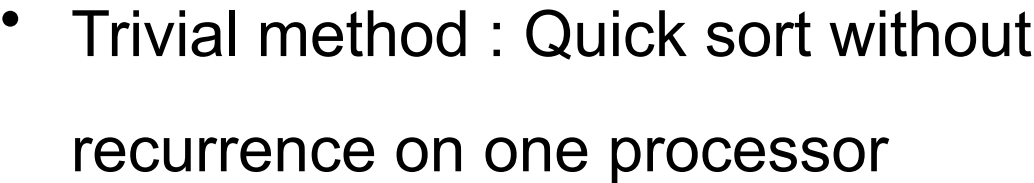


## CONTENTS

/01

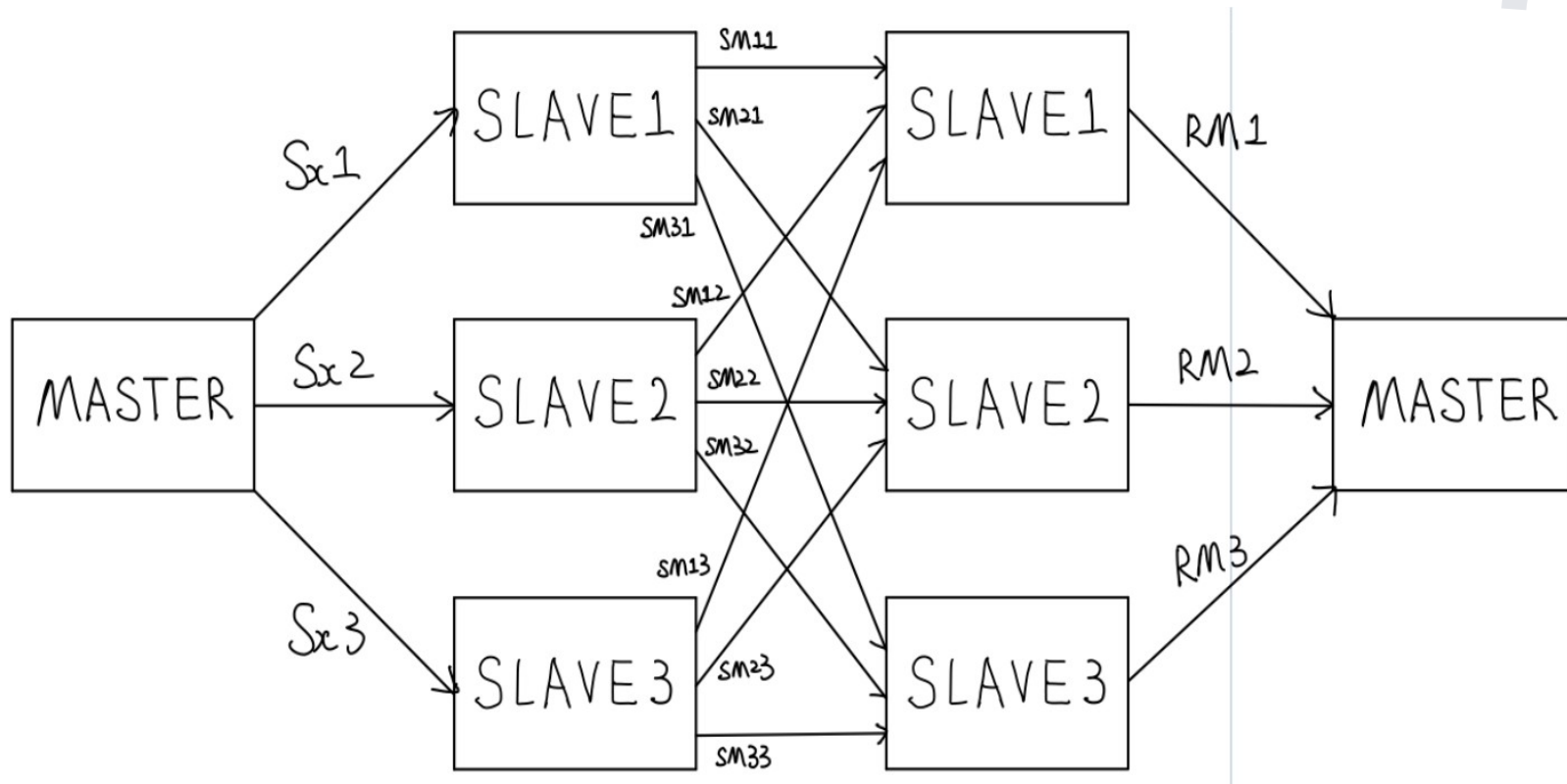
# Idea





VS

## 1.2 System diagram



### Design

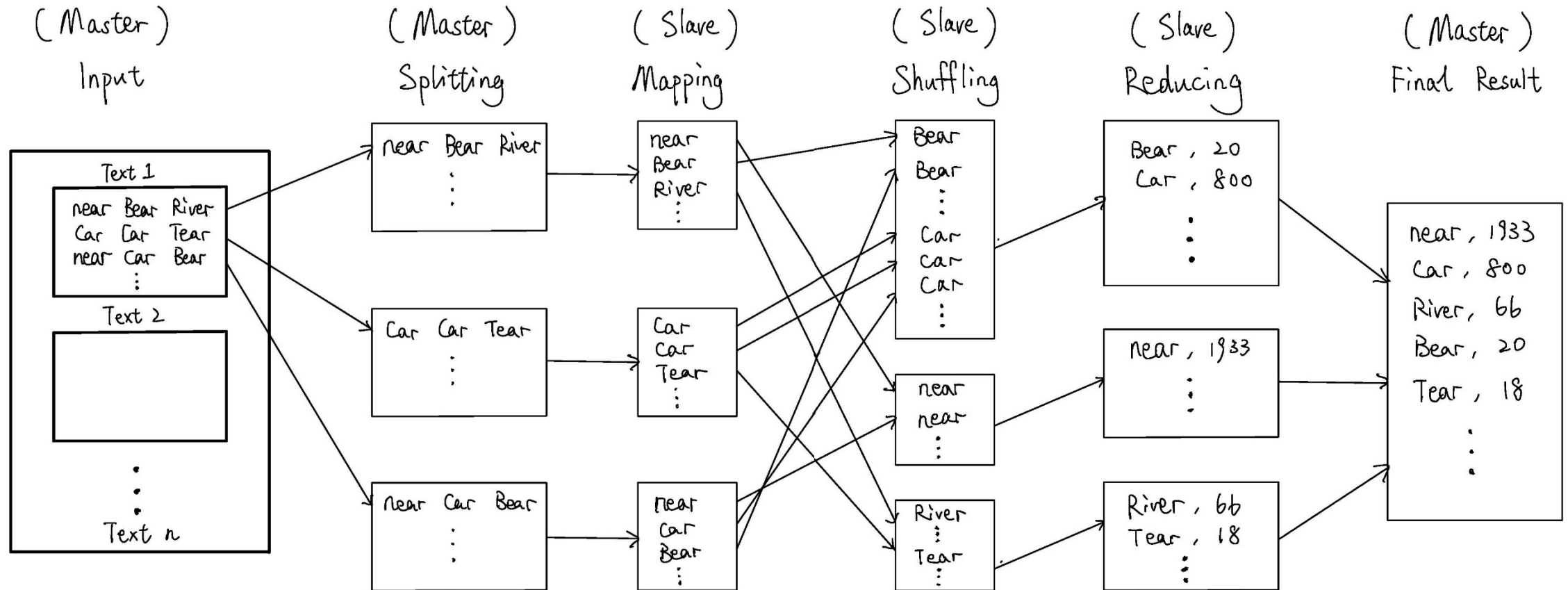
- **SLAVEs :**

- ✓ Three containers of docker
- ✓ Virtual network
- ✓ Mapping ( R & i )
- ✓ Reducing
- ✓ scp service

- **MASTER :**

- ✓ Splitting and distribution (1/3)
- ✓ Final sorting

## 1.3 Overall MapReduce word count process



/02

# Demo



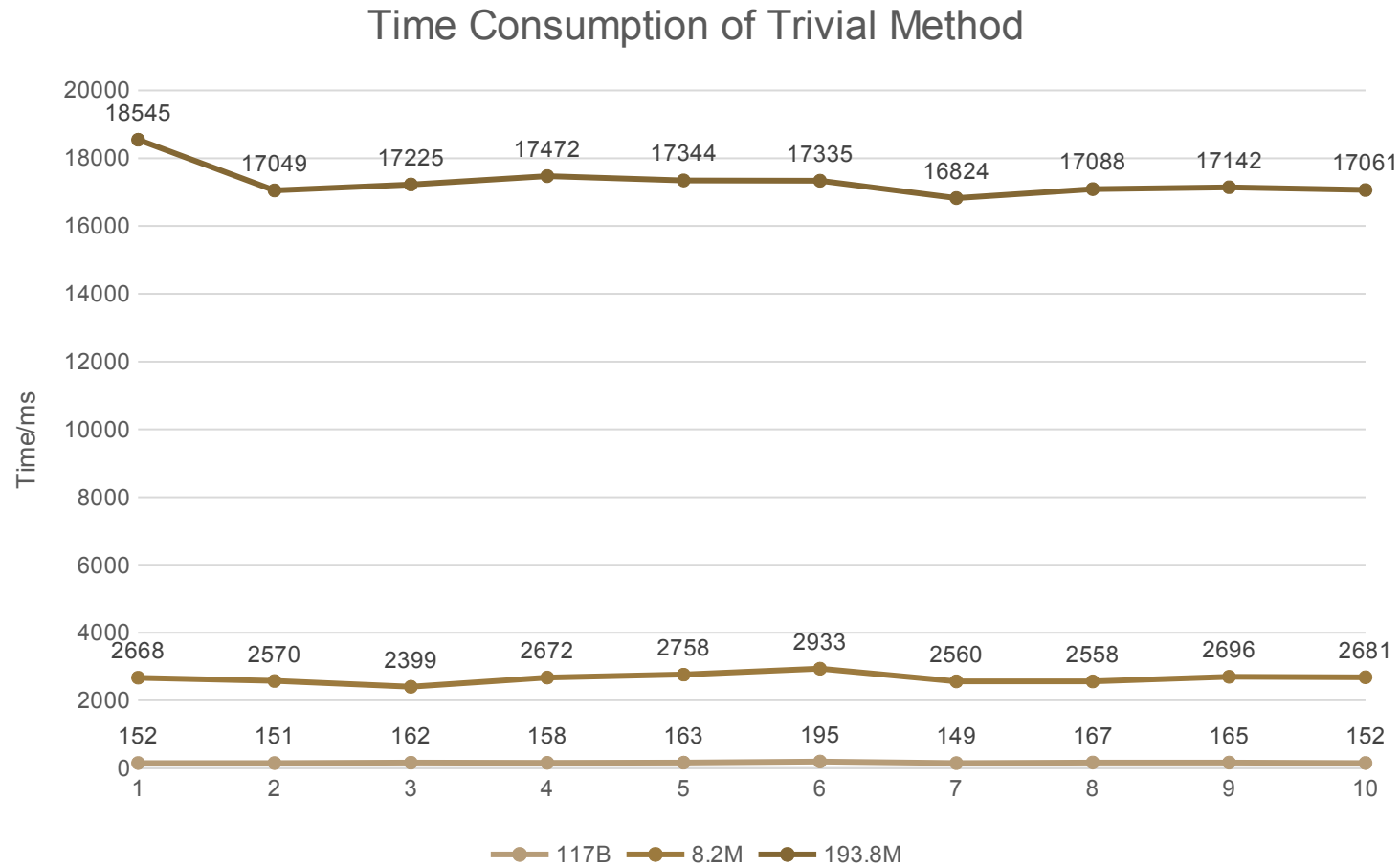
/03

# Results

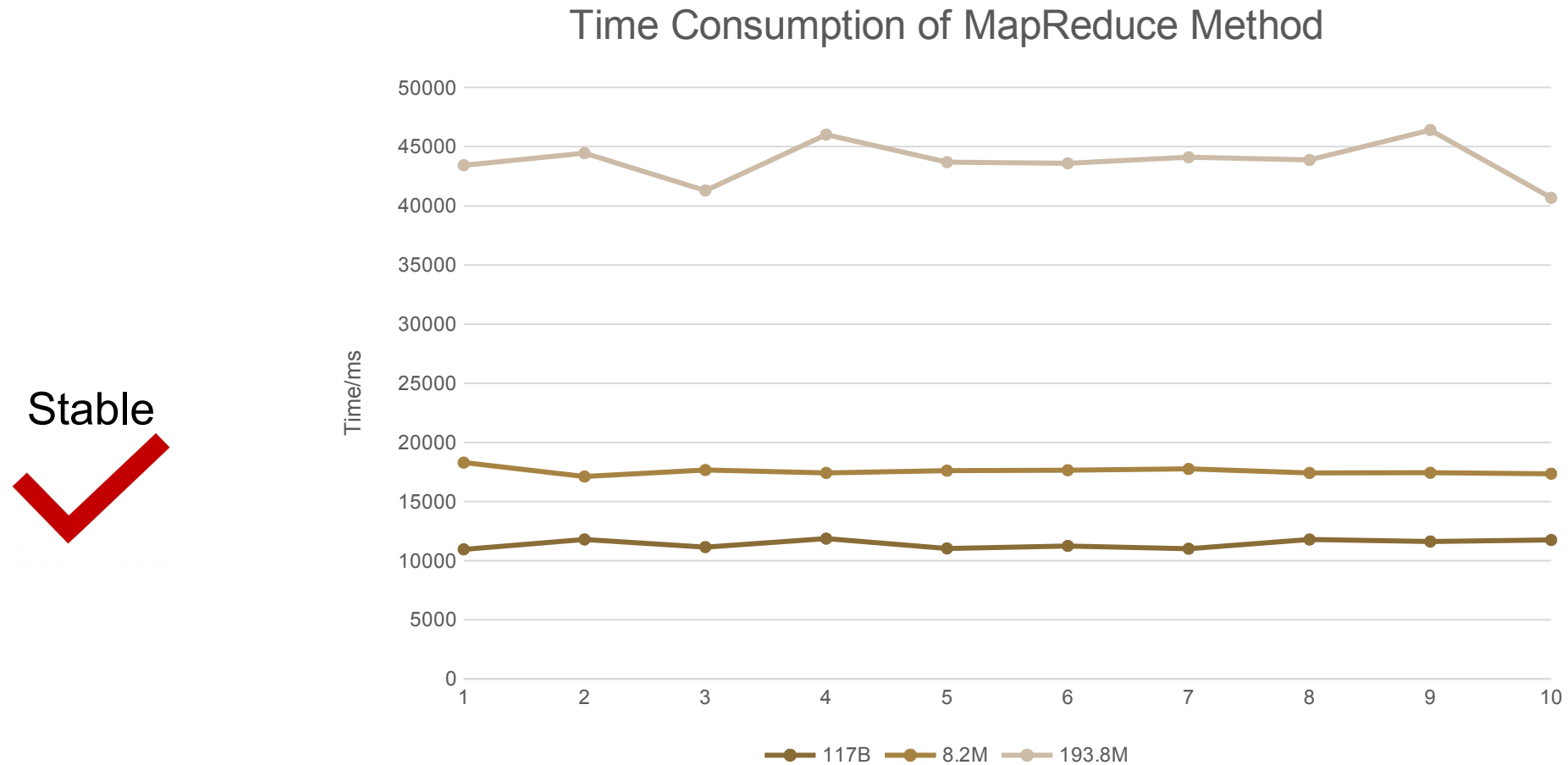


# 3.1 Time Consumption of Trivial Method

Stable  
✓



## 3.2 Time Consumption of MapReduce Method



## 3.2 Comparison of the Two Methods

File Size	117B	8.2MB	193.8MB
-----------	------	-------	---------

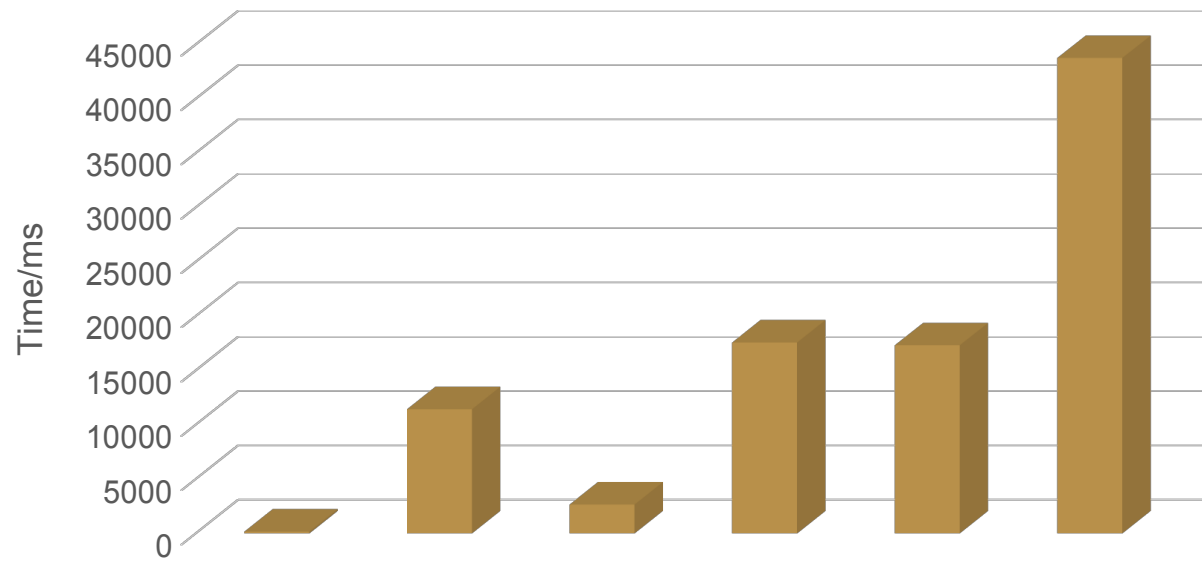
Trivial Method	161.4ms	2649.5ms	17308.5ms
----------------	---------	----------	-----------

MapReduce Method	11420.2ms	17566.4ms	43754.7ms
------------------	-----------	-----------	-----------

Why the trivial method looks better?

## 3.2 Comparison of the Two Methods

Performance of the two Methods for Files with Different Size



/04

# Improvements

## 4 Improvements

---



More slaves



Use different computers



Use larger files



We can distribute the files without reading them



In the mapping step, we divide the words by the first letter 'R' and 'I', that may not be the best choice

THANKS

The background of the image features a series of overlapping, translucent, wavy bands that flow from the top left towards the bottom right. The colors of these bands range from light lavender and pale pink to soft peach and warm, muted orange tones. The overall effect is a sense of movement and fluidity, reminiscent of smoke or liquid waves. The word 'THANKS' is written in a bold, purple, hand-drawn script font, with a thick purple underline that extends to the right.