

# Software Design Project Presentation

Team Indigo

송수민, 염재후, 임경빈

# Management

Milestone Set - 1

**Milestone #0 : Test Function for each part → (X)**

Design at least 2 unit tests before implementation, implement test function

**Milestone #1 : Setting up development environment → (O)**

Selecting development OS Scala version, JDK version Setting github directory

**Milestone #2 : Generating Dataset → (O)**

Understanding Gensort, making sample dataset

**Milestone #3 : Server and Worker communication → (O)**

Understanding gRPC Server and worker(Make document for other member)  
running test Sending and Receiving sample data

**Milestone #4 : Dataset fragmentation → (O)**

Divide single file with any size into designated sized files

# Management – Who is responsible for what

Milestone Set - 2

## **Milestone #5 : Sorting Data fragment → (O)**

Sort any single file with key, and then extend to multiple files

## **Milestone #6 : Partition Data fragment → (O)**

Label data with range given by master

## **Milestone #7 : Shuffling data fragment → (O)**

Exchange data through master so that every machine has its own labeled data

## **Milestone #8 : Merging data on each worker machine → (O)**

Sort multiple files with arbitrary size in increasing order

## **Milestone #9 : Balancing data on multiple worker machines → (X)**

By additional communication, let every machine have similar data size

# Management – Who is responsible for what

Who is responsible for what

송수민 - #1, #4, #5

염재후 - #0, #2, #3, #7

임경빈 - #3, #7, #8

#9 → give up (due to time limit)

Milestone #0 : Test Function for each part → (X)

Milestone #1 : Setting up development environment → (O)

Milestone #2 : Generating Dataset → (O)

Milestone #3 : Server and Worker communication → (O)

Milestone #4 : Dataset fragmentation → (O)

Milestone #5 : Sorting Data fragment → (O)

Milestone #6 : Partition Data fragment → (O)

Milestone #7 : Shuffling data fragment → (O)

Milestone #8 : Merging data on each worker machine → (O)

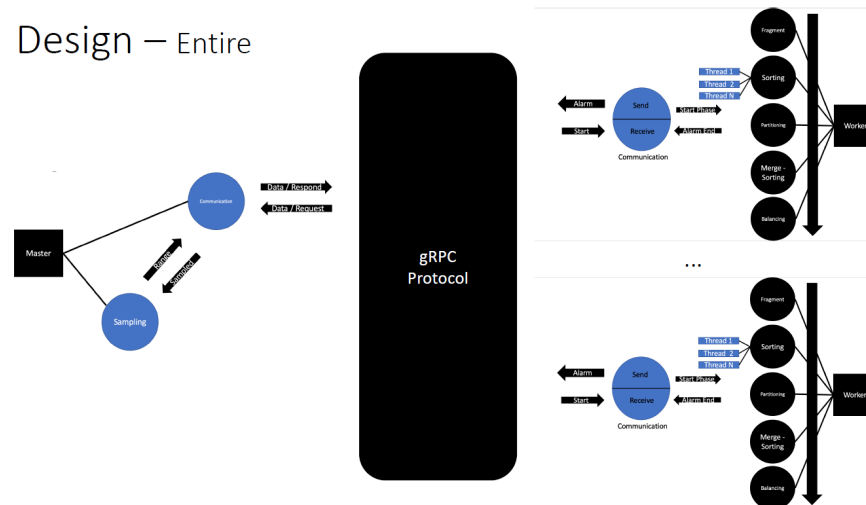
Milestone #9 : Balancing data on multiple worker machines → (X)

# Design

Anything unusual

1. all data communication drop by master
  - Consume enormous time
2. Be aware of input specification
  - Do not need Fragment phase
3. Actually, some phase can be asynchronous task
  - Initial design responded when all workers were connected

Design – Entire



# Design

Anything unusual

1. all data communication drop by master
  - shuffling are done by worker-worker communication
2. Be aware of input specification
  - Do not need Fragment phase
    - moved into merge phase.
3. Actually, some phase can be asynchronous task
  - Initial design responded when all workers were connected
    - remove waiting time

# What we learned from this project

If we were to redo...

송수민:

- focus on a short time
- make a bold turn if progress are stuck => Powerful Leader

염재후:

- start early
- elect powerful leader

임경빈:

- Although do same part, need to differentiate detailed task.
- Find implemented modules, functions