# Project Report

Team Blue

@choisuim @parksbn812

# Contents

# Plan

# Revised Plan

| Date | Subject | Todo | |
|---|---|---|---|
| Done: Connect Master to Worker, Sampling | | | |
| Dec. 1st week | Milestone #3 | Implement Pivoting<br>Implement Shuffle | Implement sorting, partitioning<br>Design Merging |
| Dec. 2nd week | Milestone #4 | Overall test | Implement Merging |
| Dec. 3rd week | Progress Presentation | | |

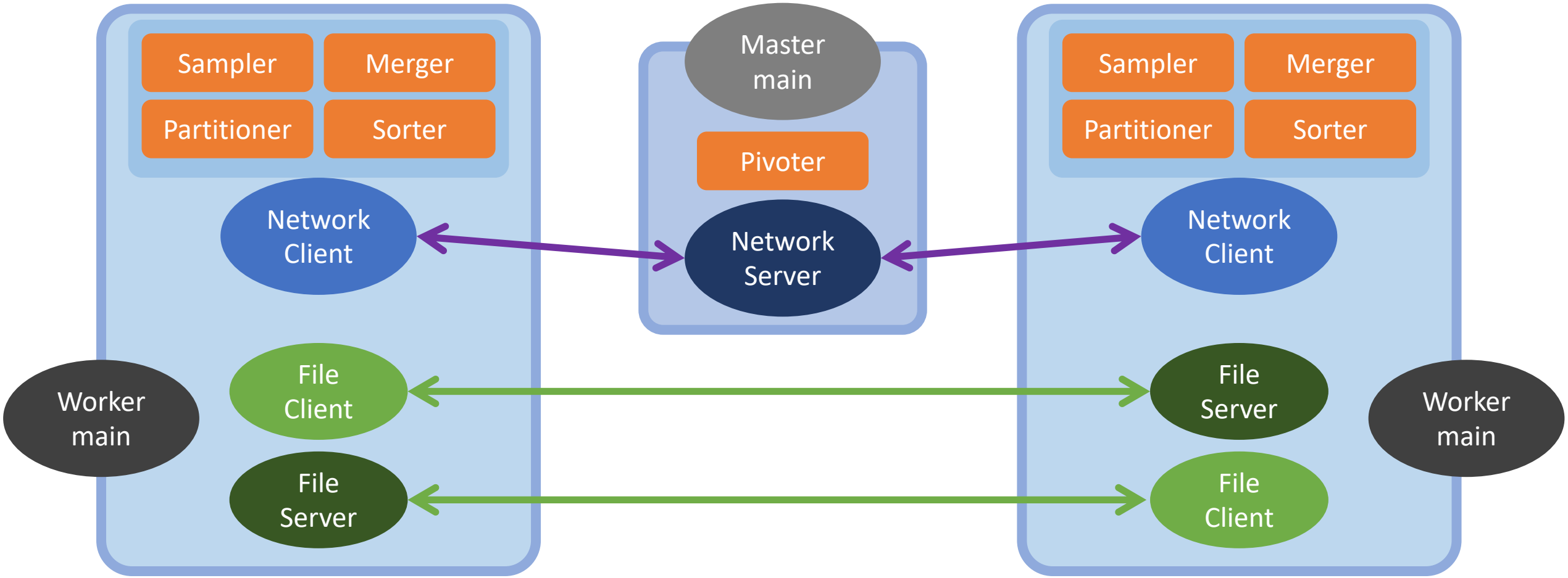
# Development

- Libraries
- Architecture & Overall communication
- Sorting Logic
- Messaging System

# Libraries

- Scalapb
  https://scalapb.github.io
- gRPC
  https://grpc.io

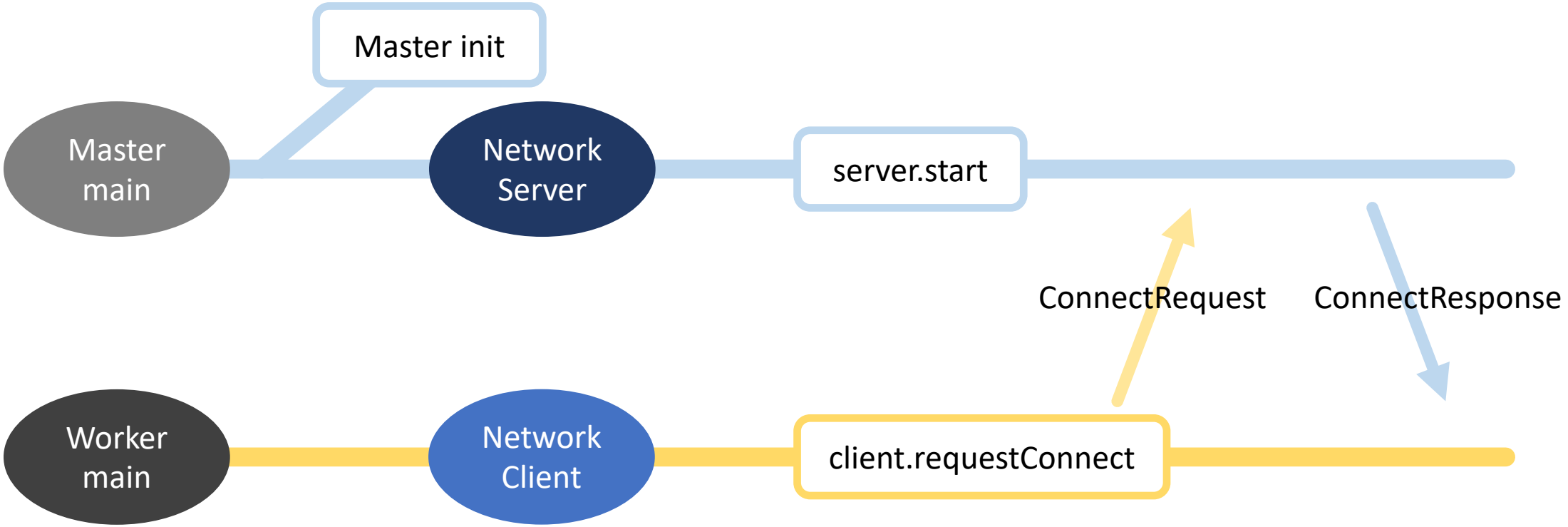
# Architecture & Overall Communication



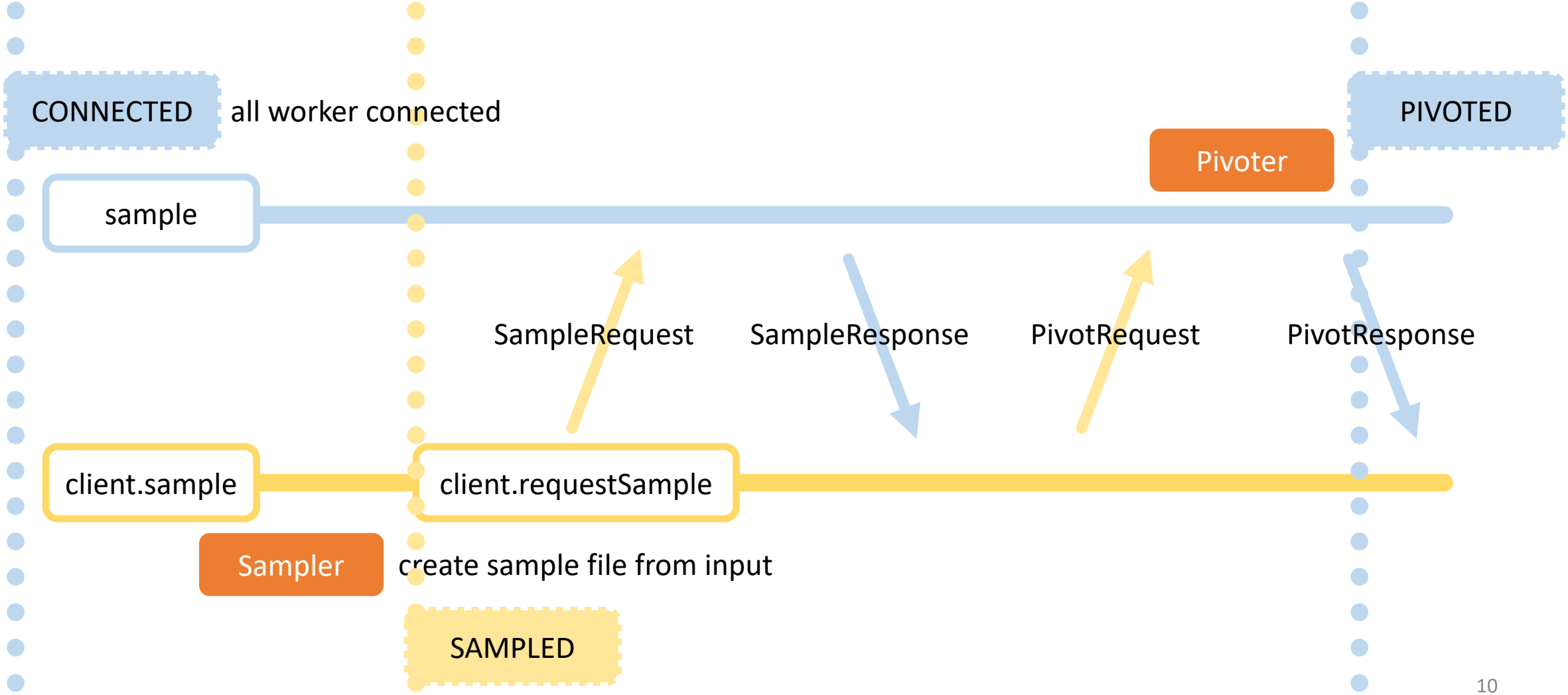
# Sorting Logics

- Get Sample from inputs
- Set pivots
- Partition each workers' inputs with pivots(ranges)
  (partition-workerId-#...)
- Shuffle each partitioned files to each workers
  (shuffle-workerId-#...)
- Merge shuffled files with each workers' subRanges
  (output-…-unsorted)
- Sort each unsorted outputs
  (output)

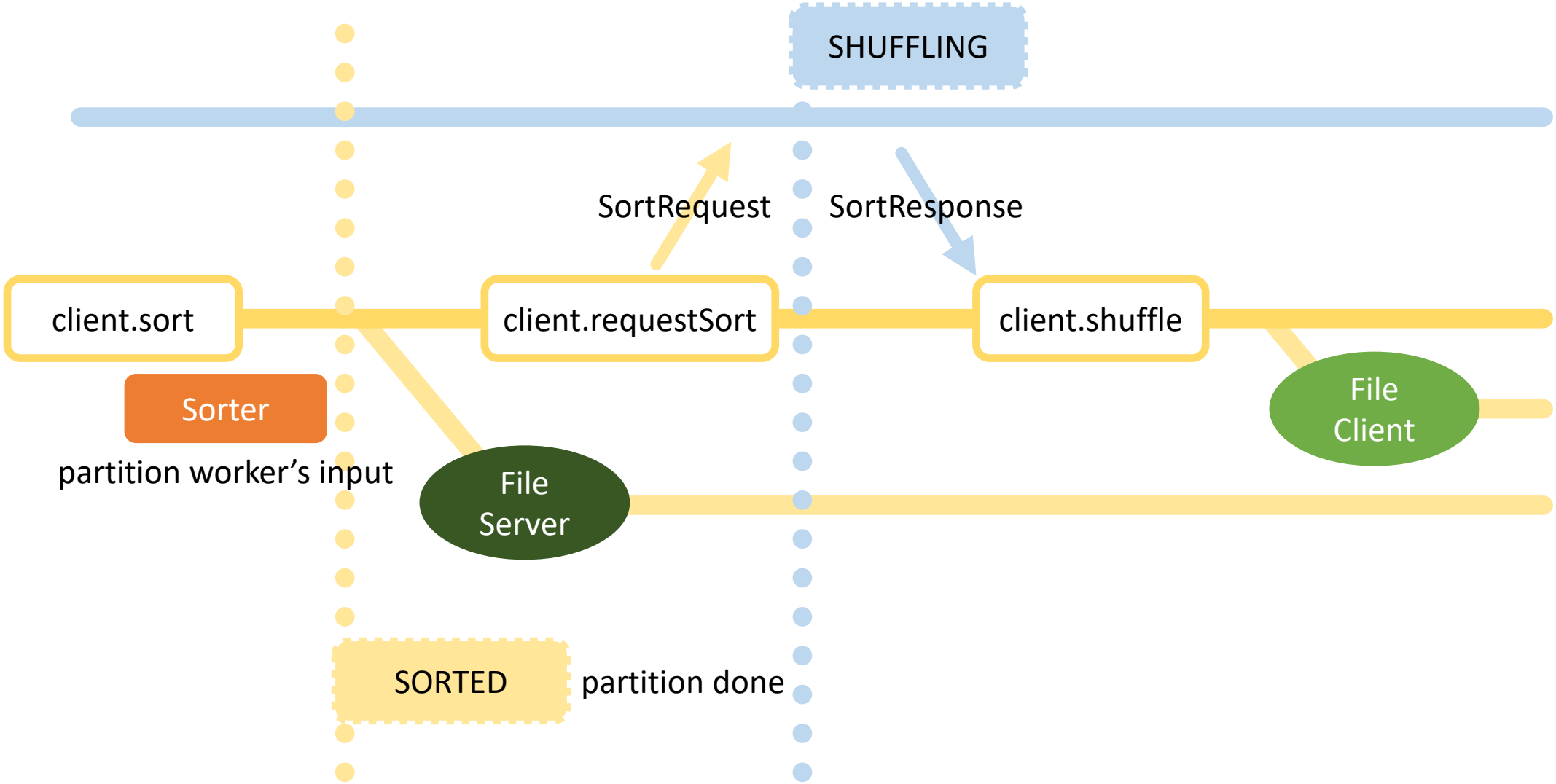
# Messaging System: Connect



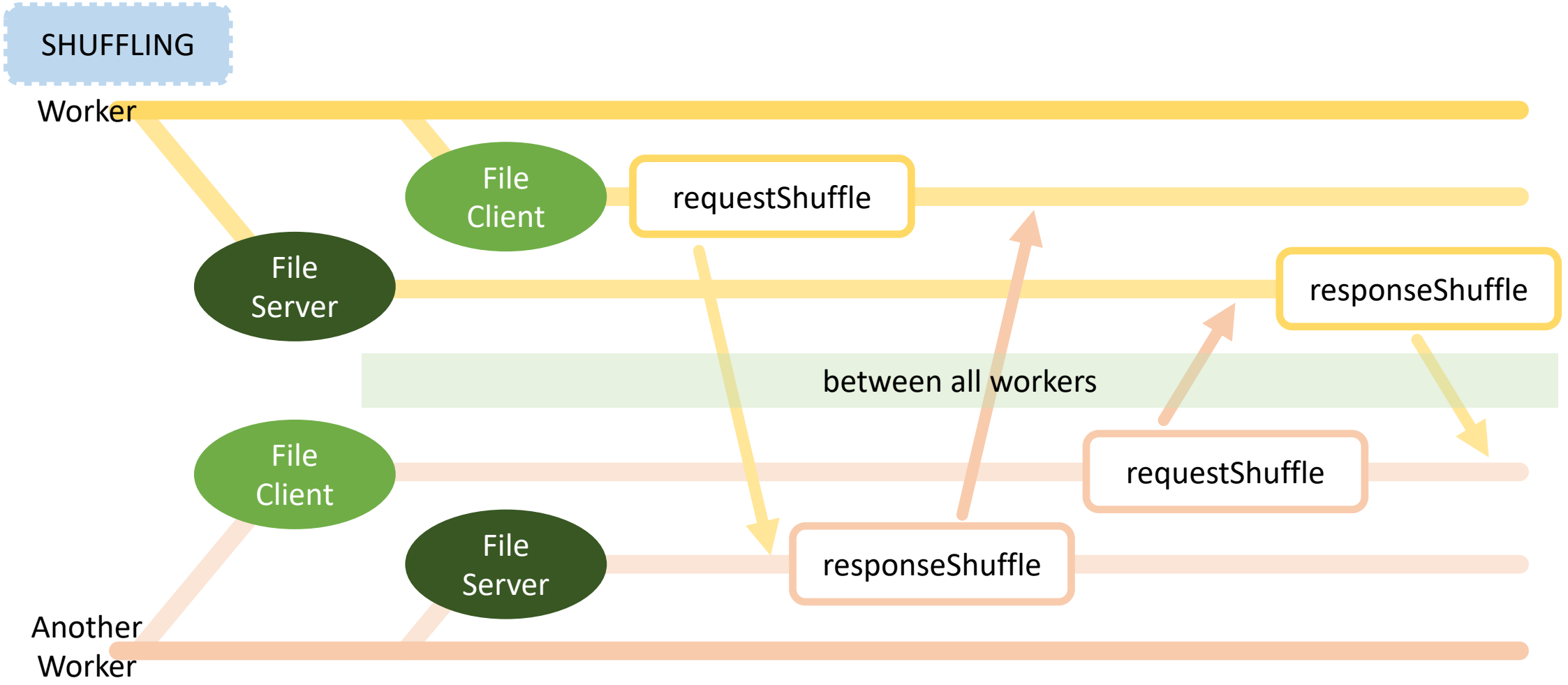
# Messaging System: Sample & Pivot



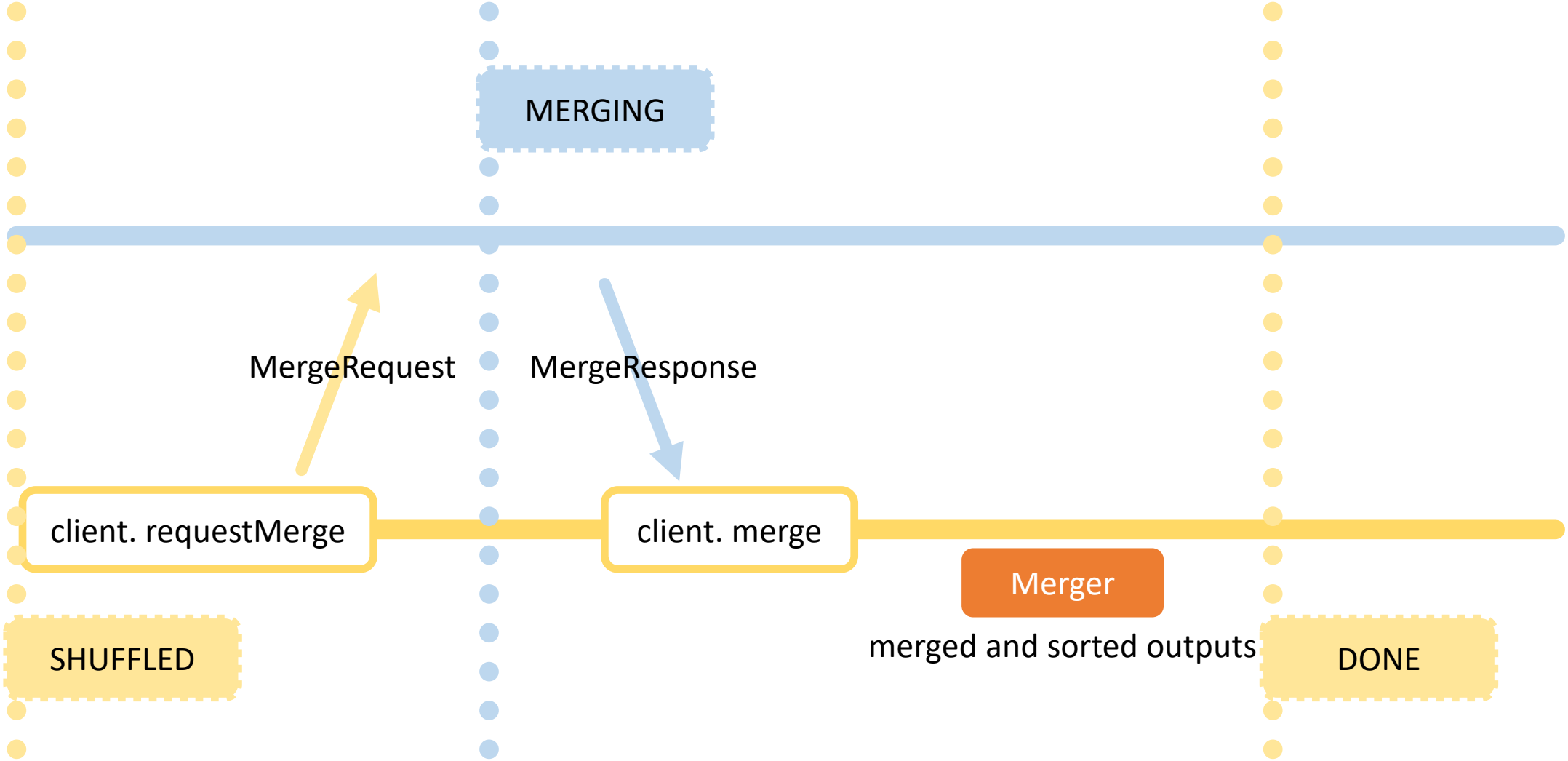
# Messaging System: Shuffle(1)



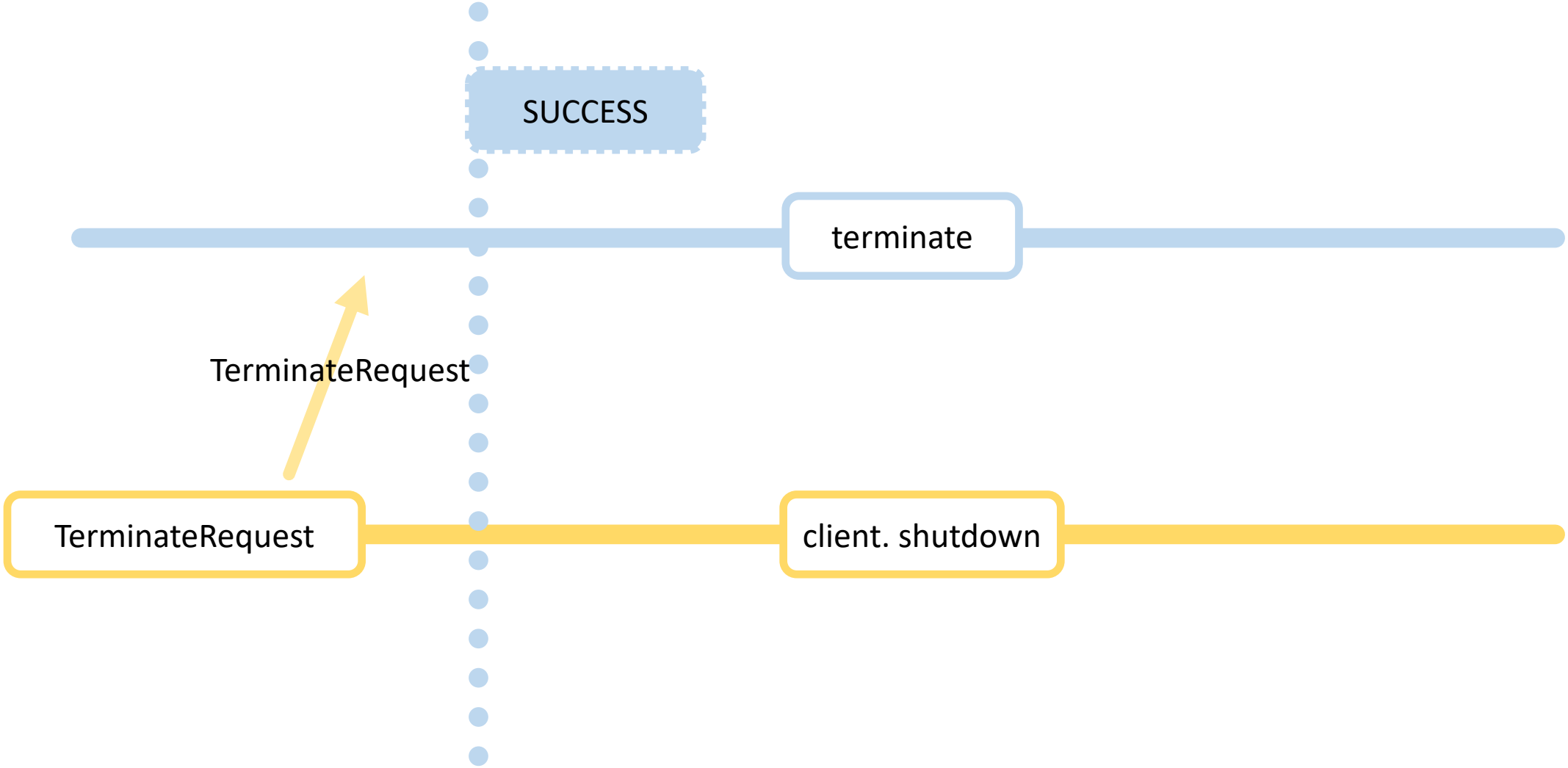
# Messaging System: Shuffle(2)



# Messaging System: Merge



# Messaging System: Terminate



# Demo Results(link)

- 1 workers with input10(306M)
- 1 workers with input300(9G)
- 3 workers with input10(306M * 3)
- 3 workers with input300(9G * 3)
- Validate
  each files with bash script

22 valsort.sh

```
@@ -0,0 +1,22 @@
+ #!/bin/bash
+
+ # Usage: . valsort.sh [output folder path] [output file count - 1]
+ # e.g. . valsort.sh /home/blue/output 299
+ # Need valsort executable file in the path ./64/valsort
+
+ shopt -s extglob
+
+ rm $1/*.sum
+
+ for outputfile in $1/output.+([0-9])!(.sum);
+ do
+   echo $outputfile
+   ./64/valsort -o $outputfile.sum $outputfile
+ done
+
+ for i in $(seq 0 $2);
+ do
+   echo $1/output.$i.sum
+   cat $1/output.$i.sum >> $1/total.sum
+ done
+ ./64/valsort -s $1/total.sum
```

# Demo Results(link)

### 1 worker with input10 (306M)

- master

```
[blue@node1 ~]$ ./master 1
192.168.10.100:8000
192.168.10.2
```

- worker

```
[blue@node2 ~]$ ./worker 192.168.10.100:8000 -I /input10 -O /home/blue/output
[blue@node2 ~]$ ls
64  output  valsort.sh  worker
[blue@node2 ~]$ ls output/
output.0  output.1  output.2  output.3  output.4  output.5  output.6  output.7  output.8  output.9
```

실행 후 output 파일이 잘 생성되었다.

```
/home/blue/output/output.0.sum
/home/blue/output/output.1.sum
/home/blue/output/output.2.sum
/home/blue/output/output.3.sum
/home/blue/output/output.4.sum
/home/blue/output/output.5.sum
/home/blue/output/output.6.sum
/home/blue/output/output.7.sum
/home/blue/output/output.8.sum
/home/blue/output/output.9.sum
Records: 3200000
Checksum: 186aa3d74df836
Duplicate keys: 0
SUCCESS - all records are in order
```

All output files are in order and the number or records is same as original.

### 1 worker with input300 (9G)

- worker

```
Records: 96000000
Checksum: 2dc6bdcfc33670c
Duplicate keys: 0
SUCCESS - all records are in order
```

```
[blue@node4 output]$ head ./output.0
    "O!uve  0000000000000000000000001228D4  77778888000022224444DDDDDDDDEEEE00000000CCCC7777DDDD
    PMd32=  0000000000000000000000003440CC1  FFFFEEEE6666CCCCBBBB999933335555DDDDDDDD777788886666
    ^3CO],  00000000000000000000000158C5C5  5555AAAA9999EEEE888822229999CCCCDDDD6666555544442222
   !&S3/]]  0000000000000000000000002145D78  8888BBBBDDDD1111CCCC55556666BBBB1111EEEEDDDD22229999
   !,=U#,9  00000000000000000000000019072E3  33332222FFFFBBBB0000FFFFAAAA666655553333DDDD3333CCCC
   !0f[ITd  0000000000000000000000003CAAB4B  9999FFFF555533337777CCCC4444BBBB7777EEEEBBBBDDDD4444
   !f6Suy2  0000000000000000000000003ABFD84  EEEE555555556666AAAA5555BBBBDDDD0000111166660000DDDD
   #%NIpq.  00000000000000000000000003B36FB9  11110000333344441111666666666AAAAAAAA00001111CCCCEEEE
   #'^cl'~  0000000000000000000000002EDC5C8  8888AAAA11114444FFFF77773333EEEE44440000FFFF99999999
   &&!epOj  0000000000000000000000001CBB42E  2222BBBB2222FFFFFFFFFFFFCCCC555566666666777700003333
[blue@node4 output]$ tail ./output.299
~~~t^lLm0&  00000000000000000000000045AF2C3  CCCC8888BBBBEEEE7777EEEEFFFF333366661111333335555CCCC
~~~uq2k#=U  0000000000000000000000002C06745  99991111DDDD222211110000FFFFEEEEFFFF33337777CCCC2222
~~~v/0&Qnm  00000000000000000000000004709701  CCCC88883333FFFF0000000000099991111FFFF777744446666
~~~yKOl:gE  0000000000000000000000002048B4F  CCCC1111444488822226666BBBB8888555577777EEEEBBBB0000
~~~yK^H.il  0000000000000000000000000463D004  44440000FFFF3333999944447777DDDDFFFFAAAA11118888DDDD
~~~yL;C'XE  0000000000000000000000005B0D211  2222EEEE3333000022221111CCCCFFFF555577774444BBBB6666
~~~zbA_ Tt  0000000000000000000000007F9F4F  BBBBCCCC666655559999FFFF8888AAAA11116666AAAABBBB0000
~~~ze0^FEg  00000000000000000000000001E06130  4444CCCCBBBB999922228888555588888CCCCFFFF000011111111
~~~}GxjWHI  0000000000000000000000000CA1345  777711118888AAAAAAAA22221111BBBB00002222BBBBCCCC2222
~~~}P;]g0g  00000000000000000000000040DA3E4  4444FFFF444466663333EEEE88888888DDDDEEEE44442222DDDD
```

All output files are in order and the number or records is same as original.

# Demo Results(link)

**3 workers with input 10 (306M * 3)**

- master

```
192.168.10.2 192.168.10.3 192.168.10.4
```

실행 후 worker를 순서대로 출력한다.

- worker 1

```
/home/blue/output//output.0.sum
/home/blue/output//output.1.sum
/home/blue/output//output.2.sum
/home/blue/output//output.3.sum
/home/blue/output//output.4.sum
/home/blue/output//output.5.sum
/home/blue/output//output.6.sum
/home/blue/output//output.7.sum
/home/blue/output//output.8.sum
/home/blue/output//output.9.sum
Records: 3266709
Checksum: 18ea758638dbc8
Duplicate keys: 2177806
SUCCESS - all records are in order
[blue@node2 ~]$
```

```
[blue@node2 ~]$ head output/output.0
    "O!uve  00000000000000000000000001228D4  77778888000022224444DDDDDDDDEEEE00000000CCCC7777DDDD
    "O!uve  00000000000000000000000001228D4  77778888000022224444DDDDDDDDEEEE00000000CCCC7777DDDD
    "O!uve  00000000000000000000000001228D4  77778888000022224444DDDDDDDDEEEE00000000CCCC7777DDDD
   ,K4a-:v  00000000000000000000000001B8132  5555EEEE888899994444FFFF1111CCCCEEEE1111EEEE6666FFFF
   ,K4a-:v  00000000000000000000000001B8132  5555EEEE888899994444FFFF1111CCCCEEEE1111EEEE6666FFFF
   ,K4a-:v  00000000000000000000000001B8132  5555EEEE888899994444FFFF1111CCCCEEEE1111EEEE6666FFFF
   =2G^9{-  00000000000000000000000000809EE  5555DDDD1111CCCC9999BBBB0000BBBBCCCCFFFFCCCC44443333
   =2G^9{-  00000000000000000000000000809EE  5555DDDD1111CCCC9999BBBB0000BBBBCCCCFFFFCCCC44443333
   =2G^9{-  00000000000000000000000000809EE  5555DDDD1111CCCC9999BBBB0000BBBBCCCCFFFFCCCC44443333
  !00VO}d0  0000000000000000000000002A143A  8888EEEEBBBBEEEE11117777CCCCCCCCEEEE4444EEEE00007777
[blue@node2 ~]$ tail output/output.9
@>.KzJa%cb  00000000000000000000000020CEE2  66663333999999990000111177778BBBB8888DDDD11117777FFFF
@>/.r2rv0R  00000000000000000000000030E020  BBBB1111AAAACCCCDDDD9999CCCCEEEE77778888FFFF66661111
@>/.r2rv0R  00000000000000000000000030E020  BBBB1111AAAACCCCDDDD9999CCCCEEEE77778888FFFF66661111
@>/.r2rv0R  00000000000000000000000030E020  BBBB1111AAAACCCCDDDD9999CCCCEEEE77778888FFFF66661111
@>/Fw?4B)v  00000000000000000000000114A21  BBBBFFFFBBBBCCCC2222BBBB77774444DDDDAAAA111122226666
@>/Fw?4B)v  00000000000000000000000114A21  BBBBFFFFBBBBCCCC2222BBBB77774444DDDDAAAA111122226666
@>/Fw?4B)v  00000000000000000000000114A21  BBBBFFFFBBBBCCCC2222BBBB77774444DDDDAAAA111122226666
@>0]%@e:r0  0000000000000000000000004E892  888877771111444477770000888833337777FFFFCCCC8888FFFF
@>0]%@e:r0  0000000000000000000000004E892  888877771111444477770000888833337777FFFFCCCC8888FFFF
@>0]%@e:r0  0000000000000000000000004E892  888877771111444477770000888833337777FFFFCCCC8888FFFF
```

# Demo Results(link)

**3 workers with input 10 (306M * 3)**

- master

```
192.168.10.2 192.168.10.3 192.168.10.4
```

실행 후 worker를 순서대로 출력한다.

- worker 1

```
/home/blue/output//output.0.sum
/home/blue/output//output.1.sum
/home/blue/output//output.2.sum
/home/blue/output//output.3.sum
/home/blue/output//output.4.sum
/home/blue/output//output.5.sum
/home/blue/output//output.6.sum
/home/blue/output//output.7.sum
/home/blue/output//output.8.sum
/home/blue/output//output.9.sum
Records: 3266709
Checksum: 18ea758638dbc8
Duplicate keys: 2177806
SUCCESS - all records are in order
[blue@node2 ~]$
```

```
[blue@node2 ~]$ head output/output.0
    "O!uve  00000000000000000000000001228D4  77778888000022224444D
    "O!uve  00000000000000000000000001228D4  77778888000022224444D
    "O!uve  00000000000000000000000001228D4  77778888000022224444D
   ,K4a-:v  00000000000000000000000001B8132  5555EEEE888899994444F
   ,K4a-:v  00000000000000000000000001B8132  5555EEEE888899994444F
   ,K4a-:v  00000000000000000000000001B8132  5555EEEE888899994444F
   =2G^9{-  00000000000000000000000000809EE  5555DDDD1111CCCC9999B
   =2G^9{-  00000000000000000000000000809EE  5555DDDD1111CCCC9999B
   =2G^9{-  00000000000000000000000000809EE  5555DDDD1111CCCC9999B
  !00VO}d0  00000000000000000000000002A143A  8888EEEEBBBBEEEE11117
[blue@node2 ~]$ tail output/output.9
@>.KzJa%cb  00000000000000000000000020CEE2  666633339999999900001
@>/.r2rv0R  00000000000000000000000030E020  BBBB1111AAAACCCCDDDD9
@>/.r2rv0R  00000000000000000000000030E020  BBBB1111AAAACCCCDDDD9
@>/.r2rv0R  00000000000000000000000030E020  BBBB1111AAAACCCCDDDD9
@>/Fw?4B)v  00000000000000000000000114A21  BBBBFFFFBBBBCCCC2222B
@>/Fw?4B)v  00000000000000000000000114A21  BBBBFFFFBBBBCCCC2222B
@>/Fw?4B)v  00000000000000000000000114A21  BBBBFFFFBBBBCCCC2222B
@>0]%@e:r0  00000000000000000000000004E892  888877771111444477770
@>0]%@e:r0  00000000000000000000000004E892  888877771111444477770
@>0]%@e:r0  00000000000000000000000004E892  888877771111444477770
```

- worker 2

```
/home/blue/output//output.0.sum
/home/blue/output//output.1.sum
/home/blue/output//output.2.sum
/home/blue/output//output.3.sum
/home/blue/output//output.4.sum
/home/blue/output//output.5.sum
/home/blue/output//output.6.sum
/home/blue/output//output.7.sum
/home/blue/output//output.8.sum
/home/blue/output//output.9.sum
Records: 3050946
Checksum: 174b590119f972
Duplicate keys: 2033964
SUCCESS - all records are in order
[blue@node3 ~]$
```

```
[blue@node3 ~]$ head output/output.0
@>0g^OQ )w  00000000000000000000000002A6520  777711113333BBBBAAAA3333777777774444BBBB666666661111
@>0g^OQ )w  00000000000000000000000002A6520  777711113333BBBBAAAA3333777777774444BBBB666666661111
@>0g^OQ )w  00000000000000000000000002A6520  777711113333BBBBAAAA3333777777774444BBBB666666661111
@>1K  e5|z  00000000000000000000000340034  00006666AAAA4444AAAA4444DDDDDDDDDDDD0000DDDD9999DDDD
@>1K  e5|z  00000000000000000000000340034  00006666AAAA4444AAAA4444DDDDDDDDDDDD0000DDDD9999DDDD
@>1K  e5|z  00000000000000000000000340034  00006666AAAA4444AAAA4444DDDDDDDDDDDD0000DDDD9999DDDD
@>1c]M{D:F  00000000000000000000000E6F49  5555BBBB4444EEEE111155553333BBBBFFFFFFFFCCCCBBBBEEEE
@>1c]M{D:F  00000000000000000000000E6F49  5555BBBB4444EEEE111155553333BBBBFFFFFFFFCCCCBBBBEEEE
@>1c]M{D:F  00000000000000000000000E6F49  5555BBBB4444EEEE111155553333BBBBFFFFFFFFCCCCBBBBEEEE
@>2aeuA-P+  00000000000000000000000001CA9DC  99998888DDDD9999CCCC22223333999988884444CCCC55555555
[blue@node3 ~]$ tail output/output.9
^QmQ-b9eB;  00000000000000000000000002DF6BB  00006666CCCC1111EEEEBBBB22220000AAAADDDD777766664444
^Qm]\,%6?\  00000000000000000000000000F5C1F  11114444333366665555333339999EEEE66665555BBBBEEEE0000
^Qm]\,%6?\  00000000000000000000000000F5C1F  11114444333366665555333339999EEEE66665555BBBBEEEE0000
^Qm]\,%6?\  00000000000000000000000000F5C1F  11114444333366665555333339999EEEE66665555BBBBEEEE0000
^Qmk~fo[\l  00000000000000000000000001F9214  CCCC555533336666666CCCC4444555577778888222233333DDDD
^Qmk~fo[\l  00000000000000000000000001F9214  CCCC555533336666666CCCC4444555577778888222233333DDDD
^Qmk~fo[\l  00000000000000000000000001F9214  CCCC555533336666666CCCC4444555577778888222233333DDDD
^Qmleg!?<#  00000000000000000000000004F751  FFFFFFFFFFFF888833338888CCCC333322228888111177776666
^Qmleg!?<#  00000000000000000000000004F751  FFFFFFFFFFFF888833338888CCCC333322228888111177776666
^Qmleg!?<#  00000000000000000000000004F751  FFFFFFFFFFFF888833338888CCCC333322228888111177776666
```

# Demo Results(link)

**3 workers with input 10 (306M * 3)**

- master

```
192.168.10.2 192.168.10.3 192.168.10.4
```

실행 후 worker를 순서대로 출력한다.

- worker 1

```
/home/blue/output//output.0.sum
/home/blue/output//output.1.sum
/home/blue/output//output.2.sum
/home/blue/output//output.3.sum
/home/blue/output//output.4.sum
/home/blue/output//output.5.sum
/home/blue/output//output.6.sum
/home/blue/output//output.7.sum
/home/blue/output//output.8.sum
/home/blue/output//output.9.sum
Records: 3266709
Checksum: 18ea758638dbc8
Duplicate keys: 2177806
SUCCESS - all records are in order
[blue@node2 ~]$
```

```
[blue@node2 ~]$ head output/output.0
    "O!uve  00000000000000000000000001228D4  77778888000022224444D
    "O!uve  00000000000000000000000001228D4  77778888000022224444D
    "O!uve  00000000000000000000000001228D4  77778888000022224444D
   ,K4a-:v  00000000000000000000000001B8132  5555EEEE888899994444F
   ,K4a-:v  00000000000000000000000001B8132  5555EEEE888899994444F
   ,K4a-:v  00000000000000000000000001B8132  5555EEEE888899994444F
   =2G^9{-  00000000000000000000000000809EE  5555DDDD1111CCCC9999B
   =2G^9{-  00000000000000000000000000809EE  5555DDDD1111CCCC9999B
   =2G^9{-  00000000000000000000000000809EE  5555DDDD1111CCCC9999B
  !00VO}d0  00000000000000000000000002A143A  8888EEEEBBBBEEEE11117
[blue@node2 ~]$ tail output/output.9
@>.KzJa%cb  00000000000000000000000020CEE2  6666333399999999000010
@>/.r2rv0R  00000000000000000000000030E020  BBBB1111AAAACCCCDDDD9
@>/.r2rv0R  00000000000000000000000030E020  BBBB1111AAAACCCCDDDD9
@>/.r2rv0R  00000000000000000000000030E020  BBBB1111AAAACCCCDDDD9
@>/Fw?4B)v  00000000000000000000000000114A21  BBBBFFFFBBBBCCCC2222B
@>/Fw?4B)v  00000000000000000000000000114A21  BBBBFFFFBBBBCCCC2222B
@>/Fw?4B)v  00000000000000000000000000114A21  BBBBFFFFBBBBCCCC2222B
@>0]%@e:r0  00000000000000000000000004E892  888877771111444477770
@>0]%@e:r0  00000000000000000000000004E892  888877771111444477770
@>0]%@e:r0  00000000000000000000000004E892  888877771111444477770
```

- worker 2

```
/home/blue/output//output.0.sum
/home/blue/output//output.1.sum
/home/blue/output//output.2.sum
/home/blue/output//output.3.sum
/home/blue/output//output.4.sum
/home/blue/output//output.5.sum
/home/blue/output//output.6.sum
/home/blue/output//output.7.sum
/home/blue/output//output.8.sum
/home/blue/output//output.9.sum
Records: 3050946
Checksum: 174b590119f972
Duplicate keys: 2033964
SUCCESS - all records are in order
[blue@node3 ~]$
```

```
[blue@node3 ~]$ head output/output.0
@>0g^OQ )w  00000000000000000000000002A6520  777711113333BBBBAAAA333377
@>0g^OQ )w  00000000000000000000000002A6520  777711113333BBBBAAAA333377
@>0g^OQ )w  00000000000000000000000002A6520  777711113333BBBBAAAA333377
@>1K  e5|z  00000000000000000000000340034  00006666AAAA4444AAAA4444DD
@>1K  e5|z  00000000000000000000000340034  00006666AAAA4444AAAA4444DD
@>1K  e5|z  00000000000000000000000340034  00006666AAAA4444AAAA4444DD
@>1c]M{D:F  00000000000000000000000E6F49  5555BBBB4444EEEE1111555533
@>1c]M{D:F  00000000000000000000000E6F49  5555BBBB4444EEEE1111555533
@>1c]M{D:F  00000000000000000000000E6F49  5555BBBB4444EEEE1111555533
@>2aeuA-P+  00000000000000000000000001CA9DC  99998888DDDD9999CCCC222233
[blue@node3 ~]$ tail output/output.9
^QmQ-b9eB;  00000000000000000000000002DF6BB  00006666CCCC1111EEEEBBBB22
^Qm]\,%6?\  00000000000000000000000000F5C1F  11114444333366665555333399
^Qm]\,%6?\  00000000000000000000000000F5C1F  11114444333366665555333399
^Qm]\,%6?\  00000000000000000000000000F5C1F  11114444333366665555333399
^Qmk~fo[\l  00000000000000000000000001F9214  CCCC55553333666666666CCCC44
^Qmk~fo[\l  00000000000000000000000001F9214  CCCC55553333666666666CCCC44
^Qmk~fo[\l  00000000000000000000000001F9214  CCCC55553333666666666CCCC44
^Qmleg!?<#  00000000000000000000000004F751  FFFFFFFFFFFF888833338888CC
^Qmleg!?<#  00000000000000000000000004F751  FFFFFFFFFFFF888833338888CC
^Qmleg!?<#  00000000000000000000000004F751  FFFFFFFFFFFF888833338888CC
```

- worker 3

```
SUCCESS - all records are in order
/home/blue/output//output.0.sum
/home/blue/output//output.1.sum
/home/blue/output//output.2.sum
/home/blue/output//output.3.sum
/home/blue/output//output.4.sum
/home/blue/output//output.5.sum
/home/blue/output//output.6.sum
/home/blue/output//output.7.sum
/home/blue/output//output.8.sum
/home/blue/output//output.9.sum
Records: 3282345
Checksum: 190a1cfe971368
Duplicate keys: 2188230
SUCCESS - all records are in order
[blue@node4 ~]$
```

```
[blue@node4 ~]$ head output/output.0
^QnIzQ=WXI  00000000000000000000000001575E9  BBBB333333335555999900000DDDD3333999999993333111EEEE
^QnIzQ=WXI  00000000000000000000000001575E9  BBBB333333335555999900000DDDD3333999999993333111EEEE
^QnIzQ=WXI  00000000000000000000000001575E9  BBBB333333335555999900000DDDD3333999999993333111EEEE
^Qnh|w-X|^  00000000000000000000000000F1B39  CCCC99994444333377778888777733339999DDDD22224444EEEE
^Qnh|w-X|^  00000000000000000000000000F1B39  CCCC99994444333377778888777733339999DDDD22224444EEEE
^Qnh|w-X|^  00000000000000000000000000F1B39  CCCC99994444333377778888777733339999DDDD22224444EEEE
^Qo;K.ZWJu  00000000000000000000000002A0E5E  BBBBAAAA2222CCCCEEEE00000000CCCCBBBBCCCC000099993333
^Qo;K.ZWJu  00000000000000000000000002A0E5E  BBBBAAAA2222CCCCEEEE00000000CCCCBBBBCCCC000099993333
^Qo;K.ZWJu  00000000000000000000000002A0E5E  BBBBAAAA2222CCCCEEEE00000000CCCCBBBBCCCC000099993333
^Qo<!{@J<,  00000000000000000000000000777DD  2222EEEE000033336666666666633330000666622222EEEEAAAA
[blue@node4 ~]$ tail output/output.9
~~}>;ArIE[  00000000000000000000000000310541  22226666FFFF1111BBBBCCCC555599993333111133330000666
~~}A3p9{^{  00000000000000000000000000132A0B  AAAA8888FFFF8888DDDDFFFF1111333300001111111111114444
~~}A3p9{^{  00000000000000000000000000132A0B  AAAA8888FFFF8888DDDDFFFF1111333300001111111111114444
~~}A3p9{^{  00000000000000000000000000132A0B  AAAA8888FFFF8888DDDDFFFF1111333300001111111111114444
~~~_jQepix  00000000000000000000000011E5D4  1111999911115555BBBB11110000222EEEE6666BBBB7777DDDD
~~~_jQepix  00000000000000000000000011E5D4  1111999911115555BBBB11110000222EEEE6666BBBB7777DDDD
~~~_jQepix  00000000000000000000000011E5D4  1111999911115555BBBB11110000222EEEE6666BBBB7777DDDD
~~~nt=ZH[N  00000000000000000000000000332A13  44441111BBBBBBBB33337777FFFF444455555555533330000CCCC
~~~nt=ZH[N  00000000000000000000000000332A13  44441111BBBBBBBB33337777FFFF444455555555533330000CCCC
~~~nt=ZH[N  00000000000000000000000000332A13  44441111BBBBBBBB33337777FFFF444455555555533330000CCCC
```

- Result summary
  - All output files in a worker are in order.
  - Also the order between workers are in order.
  - The number of records in total is same as original records(306M * 3).
  - 2/3 of records are duplicated, since the input files in a worker is same as others'.

# Demo Results(link)

**3 workers with input 300 (9G * 3)**

- worker 1

```
Records: 96848229
Checksum: 2e2e2f18f33883a
Duplicate keys: 64565486
SUCCESS - all records are in order
```

```
[blue@node2 ~]$ head ./output/output.0
    "O!uve  00000000000000000000000001228D4  777788880000222244444DDDDDDDDEEEE00000000CCCC7777DDDD
    "O!uve  00000000000000000000000001228D4  777788880000222244444DDDDDDDDEEEE00000000CCCC7777DDDD
    "O!uve  00000000000000000000000001228D4  777788880000222244444DDDDDDDDEEEE00000000CCCC7777DDDD
    PMd32=  00000000000000000000000003440CC1  FFFFEEEE6666CCCCBBBB999933335555DDDDDDDD777788886666
    PMd32=  00000000000000000000000003440CC1  FFFFEEEE6666CCCCBBBB999933335555DDDDDDDD777788886666
    PMd32=  00000000000000000000000003440CC1  FFFFEEEE6666CCCCBBBB999933335555DDDDDDDD777788886666
    ^3CO],  00000000000000000000000158C5C5  5555AAAA9999EEEE888822229999CCCCDDDD6666555544442222
    ^3CO],  00000000000000000000000158C5C5  5555AAAA9999EEEE888822229999CCCCDDDD6666555544442222
    ^3CO],  00000000000000000000000158C5C5  5555AAAA9999EEEE888822229999CCCCDDDD6666555544442222
   !&S3/]]  00000000000000000000000002145D78  8888BBBBDDDD1111CCCC55556666BBBB1111EEEEDDDD22229999
[blue@node2 ~]$ tail ./output/output.299
?z:;[q.eJH  00000000000000000000000003426479  4444DDDD0000FFFF11118888CCCC0000AAAAEEEE33330000EEEE
?z:<Mp/wTc  00000000000000000000000034D2DC0  11112222EEEEDDDD88888888877779999EEEE0000BBBB44441111
?z:<Mp/wTc  00000000000000000000000034D2DC0  11112222EEEEDDDD88888888877779999EEEE0000BBBB44441111
?z:<Mp/wTc  00000000000000000000000034D2DC0  11112222EEEEDDDD88888888877779999EEEE0000BBBB44441111
?z:="$LP|2  00000000000000000000000000A9B98D  9999666699998888DDDD7777DDDDFFFF7777EEEE7777BBBBAAAA
?z:="$LP|2  00000000000000000000000000A9B98D  9999666699998888DDDD7777DDDDFFFF7777EEEE7777BBBBAAAA
?z:="$LP|2  00000000000000000000000000A9B98D  9999666699998888DDDD7777DDDDFFFF7777EEEE7777BBBBAAAA
?z:=oD.CQ^  00000000000000000000000004EA4C  CCCCEEEEBBBBEEEEBBBBFFFF22226666333377777DDDD22225555
?z:=oD.CQ^  00000000000000000000000004EA4C  CCCCEEEEBBBBEEEEBBBBFFFF22226666333377777DDDD22225555
?z:=oD.CQ^  00000000000000000000000004EA4C  CCCCEEEEBBBBEEEEBBBBFFFF22226666333377777DDDD22225555
```

# Demo Results(link)

**3 workers with input 300 (9G * 3)**

- worker 1

```
Records: 96848229
Checksum: 2e2e2f18f3
Duplicate keys: 6456
SUCCESS - all record
```

```
[blue@node2 ~]$ head ./output/output
    "O!uve  0000000000000000000000000
    "O!uve  0000000000000000000000000
    "O!uve  0000000000000000000000000
    PMd32=  0000000000000000000000000
    PMd32=  0000000000000000000000000
    PMd32=  0000000000000000000000000
    ^3CO],  0000000000000000000000000
    ^3CO],  0000000000000000000000000
    ^3CO],  0000000000000000000000000
   !&S3/]]  0000000000000000000000000
[blue@node2 ~]$ tail ./output/output
?z:;[q.eJH  0000000000000000000000000
?z:<Mp/wTc  0000000000000000000000000
?z:<Mp/wTc  0000000000000000000000000
?z:<Mp/wTc  0000000000000000000000000
?z:="$LP|2  0000000000000000000000000
?z:="$LP|2  0000000000000000000000000
?z:="$LP|2  0000000000000000000000000
?z:=oD.CQ^  0000000000000000000000000
?z:=oD.CQ^  0000000000000000000000000
?z:=oD.CQ^  0000000000000000000000000
```

- worker 2

```
Records: 90533742
Checksum: 2b2b65653dd5b93
Duplicate keys: 60355828
SUCCESS - all records are in order
```

```
[blue@node3 ~]$ head ./output/output.0
?z:?,qPl1N  00000000000000000000000017F4C15  555599990000444411111111100008888FFFFCCCCEEEE77772222
?z:?,qPl1N  00000000000000000000000017F4C15  555599990000444411111111100008888FFFFCCCCEEEE77772222
?z:?,qPl1N  00000000000000000000000017F4C15  555599990000444411111111100008888FFFFCCCCEEEE77772222
?z:?msGvM{  00000000000000000000000032747B6  1111CCCC6666AAAA88882222EEEEAAAA77773333DDDD9999BBBB
?z:?msGvM{  00000000000000000000000032747B6  1111CCCC6666AAAA88882222EEEEAAAA77773333DDDD9999BBBB
?z:?msGvM{  00000000000000000000000032747B6  1111CCCC6666AAAA88882222EEEEAAAA77773333DDDD9999BBBB
?z:B&;5>^%  00000000000000000000000001EF0F5E  EEEEEEEE666666666666999933333333777733333333399993333
?z:B&;5>^%  00000000000000000000000001EF0F5E  EEEEEEEE666666666666999933333333777733333333399993333
?z:B&;5>^%  00000000000000000000000001EF0F5E  EEEEEEEE666666666666999933333333777733333333399993333
?z:B=o`N"@  00000000000000000000000D146D5  44440000DDDD0000AAAACCCCDDDDDDDDBBBB0000DDDDBBBB2222
[blue@node3 ~]$ tail ./output/output.299
]lVR9UXsFt  00000000000000000000000001B940A9  1111EEEEDDDDFFFF5555DDDDCCCC000033335555EEEE5555EEEE
]lVRPlzzkj  00000000000000000000000042CA0D6  66660000EEEE1111999955556666FFFFBBBB0000AAAAFFFFBBBB
]lVRPlzzkj  00000000000000000000000042CA0D6  66660000EEEE1111999955556666FFFFBBBB0000AAAAFFFFBBBB
]lVRPlzzkj  00000000000000000000000042CA0D6  66660000EEEE1111999955556666FFFFBBBB0000AAAAFFFFBBBB
]lVTQ~x;q_  00000000000000000000000053CB189  5555FFFF666633333333BBBB888833336666AAAACCCC7777EEEE
]lVTQ~x;q_  00000000000000000000000053CB189  5555FFFF666633333333BBBB888833336666AAAACCCC7777EEEE
]lVTQ~x;q_  00000000000000000000000053CB189  5555FFFF666633333333BBBB888833336666AAAACCCC7777EEEE
]lVU-3xz$D  00000000000000000000000004FE61  BBBBEEEE22225555FFFF3333DDDD3333FFFF00006666EEEE6666
]lVU-3xz$D  00000000000000000000000004FE61  BBBBEEEE22225555FFFF3333DDDD3333FFFF00006666EEEE6666
]lVU-3xz$D  00000000000000000000000004FE61  BBBBEEEE22225555FFFF3333DDDD3333FFFF00006666EEEE6666
```

# Demo Results(link)

**3 workers with input 300 (9G * 3)**

- worker 1

```
Records: 96848229
Checksum: 2e2e2f18f3
Duplicate keys: 6456
SUCCESS - all record
```

```
[blue@node2 ~]$ head ./output/output
    "O!uve  0000000000000000000000000
    "O!uve  0000000000000000000000000
    "O!uve  0000000000000000000000000
    PMd32=  0000000000000000000000000
    PMd32=  0000000000000000000000000
    PMd32=  0000000000000000000000000
    ^3CO],  0000000000000000000000000
    ^3CO],  0000000000000000000000000
    ^3CO],  0000000000000000000000000
   !&S3/]]  0000000000000000000000000
[blue@node2 ~]$ tail ./output/output
?z:;[q.eJH  0000000000000000000000000
?z:<Mp/wTc  0000000000000000000000000
?z:<Mp/wTc  0000000000000000000000000
?z:<Mp/wTc  0000000000000000000000000
?z:="$LP|2  0000000000000000000000000
?z:="$LP|2  0000000000000000000000000
?z:="$LP|2  0000000000000000000000000
?z:=oD.CQ^  0000000000000000000000000
?z:=oD.CQ^  0000000000000000000000000
?z:=oD.CQ^  0000000000000000000000000
```

- worker 2

```
Records: 90533742
Checksum: 2b2b65653dd5b
Duplicate keys: 6035582
SUCCESS - all records a
```

```
[blue@node3 ~]$ head ./output/output.0
?z:?,qPl1N  00000000000000000000000017F4
?z:?,qPl1N  00000000000000000000000017F4
?z:?,qPl1N  00000000000000000000000017F4
?z:?msGvM{  0000000000000000000000032747
?z:?msGvM{  0000000000000000000000032747
?z:?msGvM{  0000000000000000000000032747
?z:B&;5>^%  00000000000000000000000001EF0F
?z:B&;5>^%  00000000000000000000000001EF0F
?z:B&;5>^%  00000000000000000000000001EF0F
?z:B=o`N"@  00000000000000000000000D146
[blue@node3 ~]$ tail ./output/output.299
]lVR9UXsFt  000000000000000000000001B940
]lVRPlzzkj  0000000000000000000000042CA0
]lVRPlzzkj  0000000000000000000000042CA0
]lVRPlzzkj  0000000000000000000000042CA0
]lVTQ~x;q_  00000000000000000000000053CB1
]lVTQ~x;q_  00000000000000000000000053CB1
]lVTQ~x;q_  00000000000000000000000053CB1
]lVU-3xz$D  00000000000000000000000004FE
]lVU-3xz$D  00000000000000000000000004FE
]lVU-3xz$D  00000000000000000000000004FE
```

- worker 3

```
Records: 100618029
Checksum: 2ffaa4f11895157
Duplicate keys: 67078686
SUCCESS - all records are in order
```

```
[blue@node4 output]$ head ./output.0
]lVVTdP/qu  0000000000000000000000000FA3E6E  EEEE777711118888222222222DDDD66665555BBBB1111CCCC3333
]lVVTdP/qu  0000000000000000000000000FA3E6E  EEEE777711118888222222222DDDD66665555BBBB1111CCCC3333
]lVVTdP/qu  0000000000000000000000000FA3E6E  EEEE777711118888222222222DDDD66665555BBBB1111CCCC3333
]lVWlfr4>L  00000000000000000000000003A3A7E  BBBB55556666DDDDEEEE0000333322222222BBBBAAAAFFFF3333
]lVWlfr4>L  00000000000000000000000003A3A7E  BBBB55556666DDDDEEEE0000333322222222BBBBAAAAFFFF3333
]lVWlfr4>L  00000000000000000000000003A3A7E  BBBB55556666DDDDEEEE0000333322222222BBBBAAAAFFFF3333
]lVX)C!r+Y  00000000000000000000000003FFB8FD  77770000999922220000111199998888EEEE2222CCCCCCCCAAAA
]lVX)C!r+Y  00000000000000000000000003FFB8FD  77770000999922220000111199998888EEEE2222CCCCCCCCAAAA
]lVX)C!r+Y  00000000000000000000000003FFB8FD  77770000999922220000111199998888EEEE2222CCCCCCCCAAAA
]lVY"+GPff  0000000000000000000000059C66B1  55557777BBBBCCCCEEEEFFFF55558888AAAACCCC111111116666
[blue@node4 output]$ tail ./output.299
~~~zbA_ Tt  0000000000000000000000007F9F4F  BBBBCCCC666655559999FFFF8888AAAA11116666AAAABBBB0000
~~~zeO^FEg  0000000000000000000000001E06130  4444CCCCBBBB999922228888555588CCCCFFFF000011111111
~~~zeO^FEg  0000000000000000000000001E06130  4444CCCCBBBB999922228888555588CCCCFFFF000011111111
~~~zeO^FEg  0000000000000000000000001E06130  4444CCCCBBBB999922228888555588CCCCFFFF000011111111
~~~}GxjWHI  0000000000000000000000000CA1345  777711118888AAAAAAAA22221111BBBB00002222BBBBCCCC2222
~~~}GxjWHI  0000000000000000000000000CA1345  777711118888AAAAAAAA22221111BBBB00002222BBBBCCCC2222
~~~}GxjWHI  0000000000000000000000000CA1345  777711118888AAAAAAAA22221111BBBB00002222BBBBCCCC2222
~~~}P;]g0g  0000000000000000000000040DA3E4  4444FFFF444466663333EEEE88888888DDDDEEEE44442222DDDD
~~~}P;]g0g  0000000000000000000000040DA3E4  4444FFFF444466663333EEEE88888888DDDDEEEE44442222DDDD
~~~}P;]g0g  0000000000000000000000040DA3E4  4444FFFF444466663333EEEE88888888DDDDEEEE44442222DDDD
```

# Lessons(1)

- Don’t be optimistic
- When planning..
  - Detailed plan is needed
  - More investigation is needed
  - Implemented separately -> Some waiting time when connect each modules
- Surgeon Team: A person designed module
- Face-to-face is much better than online meeting
- Testing
  - TDD: Applied well at first, but gradually diminished -> result: hard to debug
  - Test with Big files

# Lessons(2)

- Load per period
  - Network: almost constant. (small failure by Netty)
  - Sorting: Exponential (implemented almost at once)
  - Integration: followed sorting
- Others
  - Lack of (physical) time
  - Not used to Scala

# Thank you

:->