

MAPREDUCE

WHAT IS MAPREDUCE?

- Distributed Programming model for processing large data sets
- Conceived at Google
- Can be implemented in any programming language
- MapReduce is NOT a programming language
- Hadoop implements MapReduce
- MapReduce System (Hadoop) - Manage communications, data transfers, parallel execution across distributed servers



MAPREDUCE

SFO



SFO 5
SFO 3
SFO 4
SFO 1

SJOSE



SJSOE 2
SJSOE 6

LA



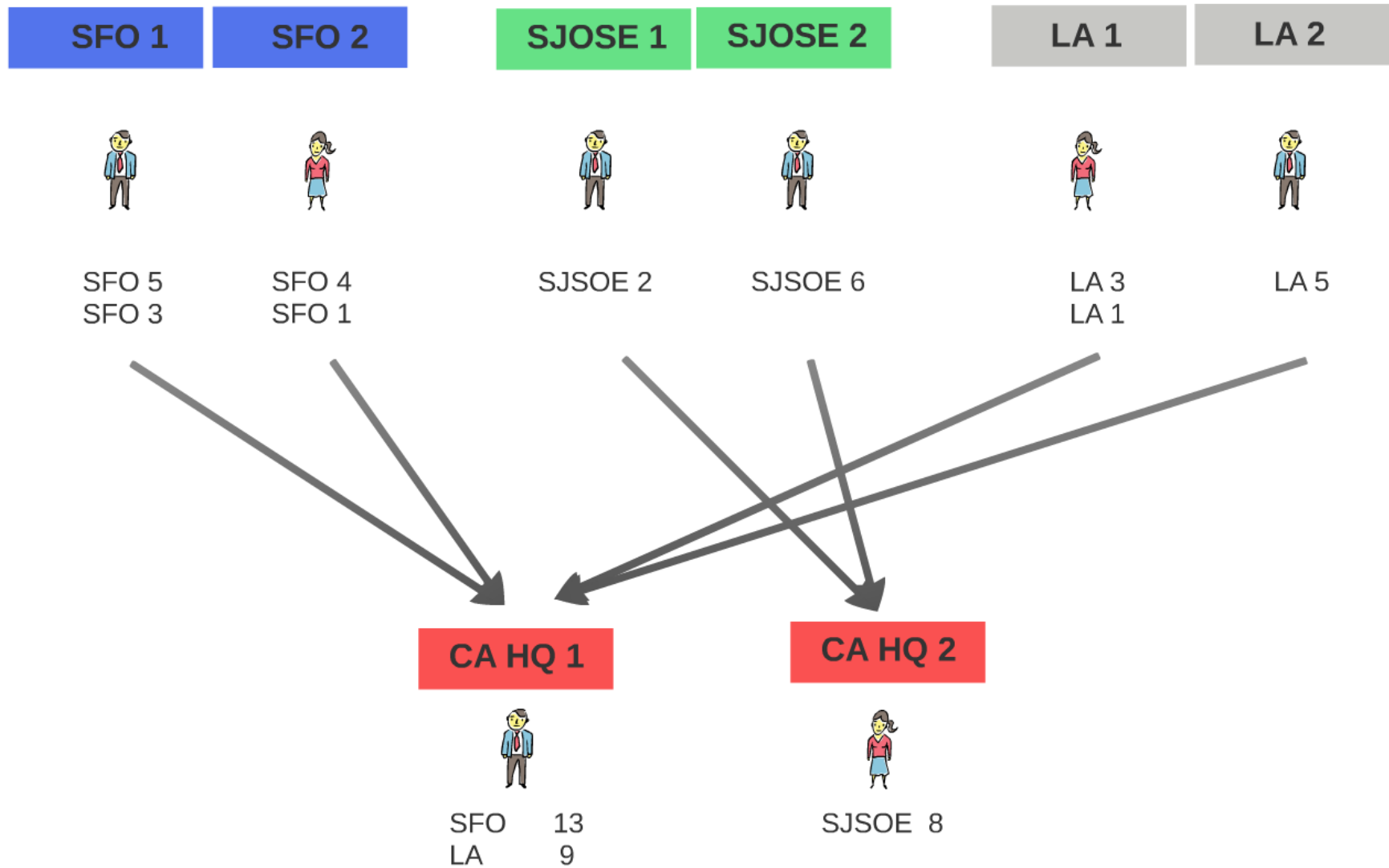
LA 3
LA 1
LA 5



CA HQ



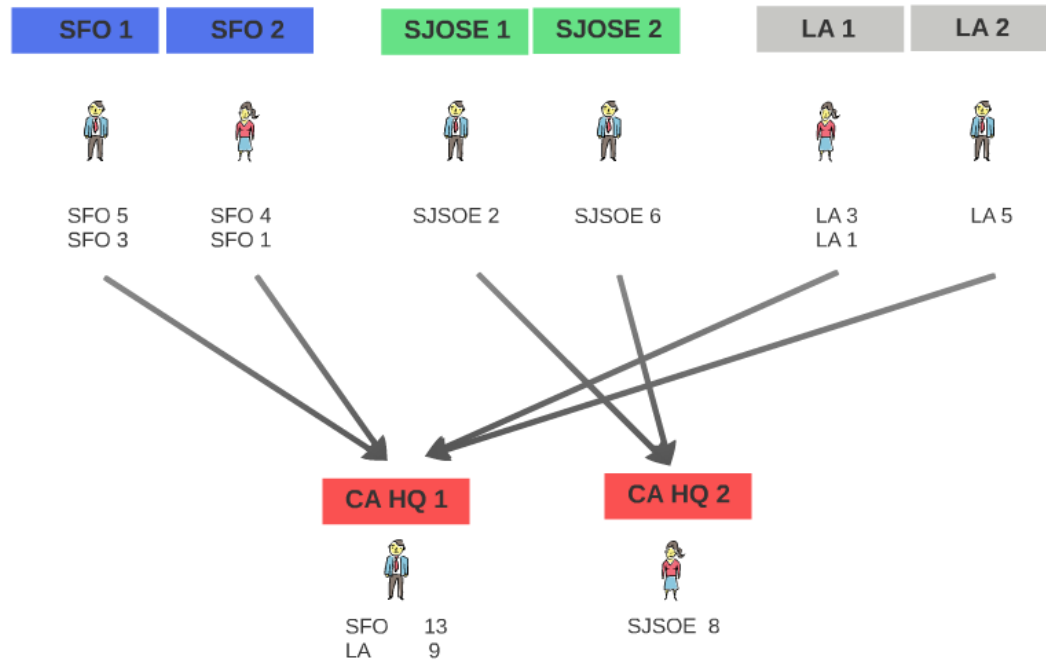
SFO	13
SJSOE	8
LA	9



MAP PHASE

SHUFFLE PHASE

REDUCE PHASE



Input Splits

Mappers

Key-Value Pairs

Reducers

Result

WHAT IS MAPREDUCE?

- Distributed Programming model for processing large data sets
- Conceived at Google
- Can be implemented in any programming language
- MapReduce is NOT a programming language
- Hadoop implements MapReduce
- MapReduce System (Hadoop) - Manage communications, data transfers, parallel execution across distributed servers

