CENG334
OPERATING SYSTEMS
THE1 RECITATION

By Erbil Yakışkan

Outline

- Homework Text
- Example Case
- Questions

Homework Text

MapReduce

- A system designed by Google Inc. in 2004.
- Handles data flow for distributed applications.
- 2 important operations:
  - Map(): Divide data into chunks and apply the same function to them.
  - Reduce(): Retrieve results from Map() and combine them to produce the overall output.

MapReduce

- For the purposes of this THE:
  - There will be the same number of mappers and reducers.
  - Mappers will only send data to their corresponding reducer.

What you will do?

- You will implement 2 models:
  - Map Model
    - There will be N mappers.
    - Mappers take input from the parent.
    - Each mapper prints its result to tty.
  - MapReduce Model
    - Extension of Map model.
    - There will be N reducers, reducer K will take input from mapper K.
    - Reducer K will take input from Recuder K-1, redirected to its stderr.

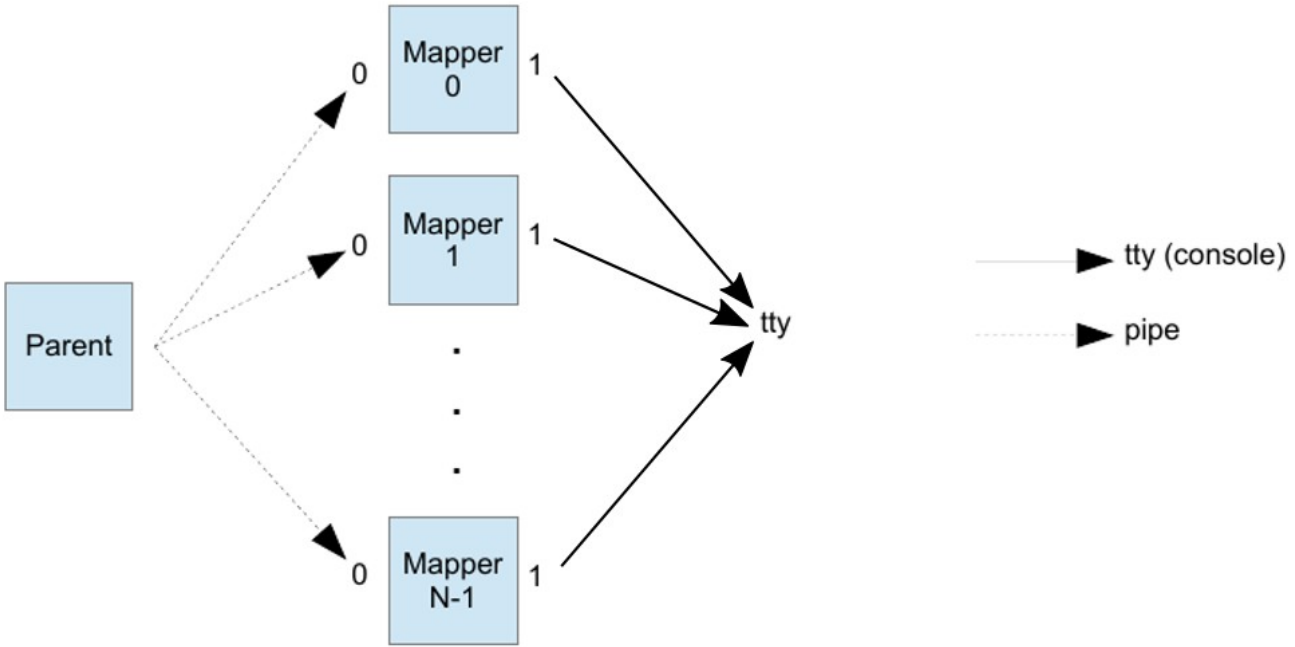
Execution

- Your code will be executed with the following command:
  - ./mapreduce N mapper_proc [reducer_proc]
  - N : number of mappers and reducers.
  - mapper_proc: path to mapper process.
  - reducer_proc: path to reducer process.
- If reducer_proc is not provided build Map model.

General Operations

- Create pipes.(pipe())
- Create child processes.(fork())
- Duplicate pipes to specified file descriptors.(dup(),dup2())
- Execute mappers and reducers.(exec family)
- Feed mappers with the given inputs.(write())
- Wait until all child processes terminate.(wait family)

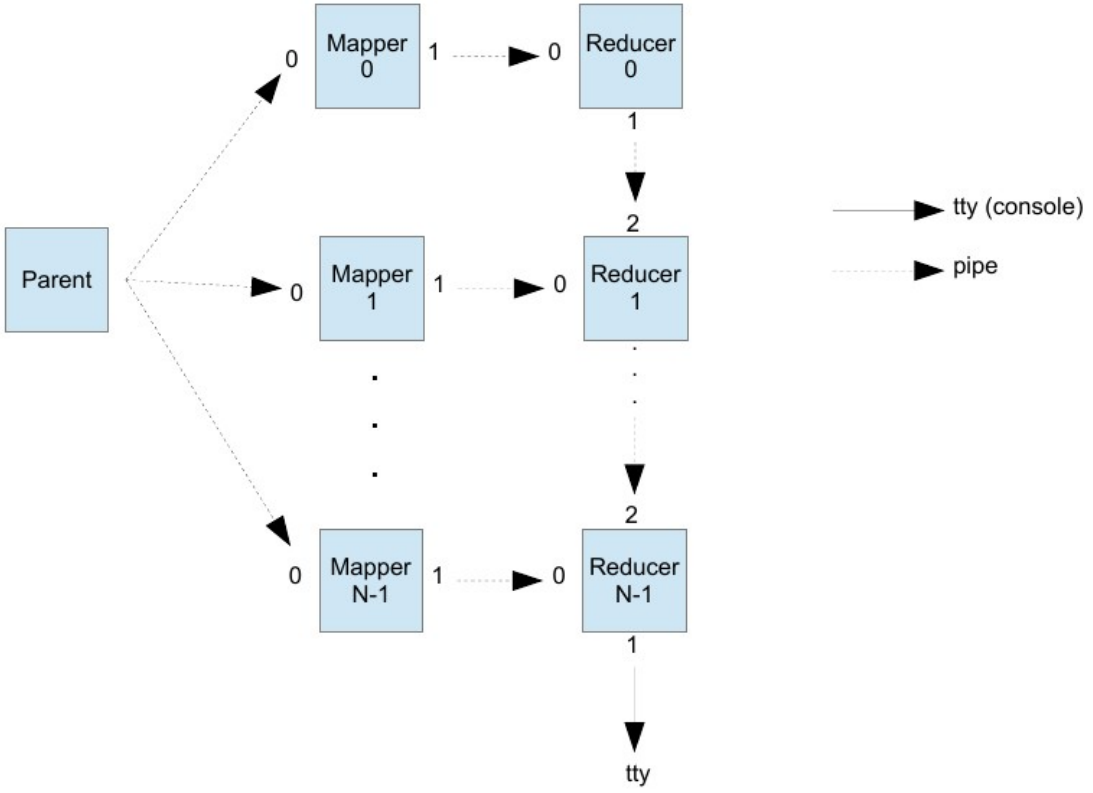
Overview of Map Model



Overview of Map Model

- Your program is the parent.
- Parent process creates the pipes, then forks the mappers.
- The parent process distributes the input line by line
- Line i of the input should be sent to Mapper (i mod N)-1.

Overview of MapReduce Model



Overview of MapReduce Model

- Map part is built in the same way as described.
- Pipes between mapper-reducer and reducer-reducer should be created.
- Reducers are also forked from parent.
- Stdout of mapper k should be connected to stdin of reducer k.
- Stdout of reducer k should be connected to stderr of reducer k+1.

General Advice

- Close all pipe ends that are not used.
- Close parent-mapper pipes after EOF is read in parent.
- No need to write separate Map model for MapReduce.
- Try to start working as soon as possible.

Good luck with your homework!
:)