

Distributed Computing System for Large-scale 3D Reconstruction

Tianwei Huang
Advisor: Prof. Long Quan



Introduction:

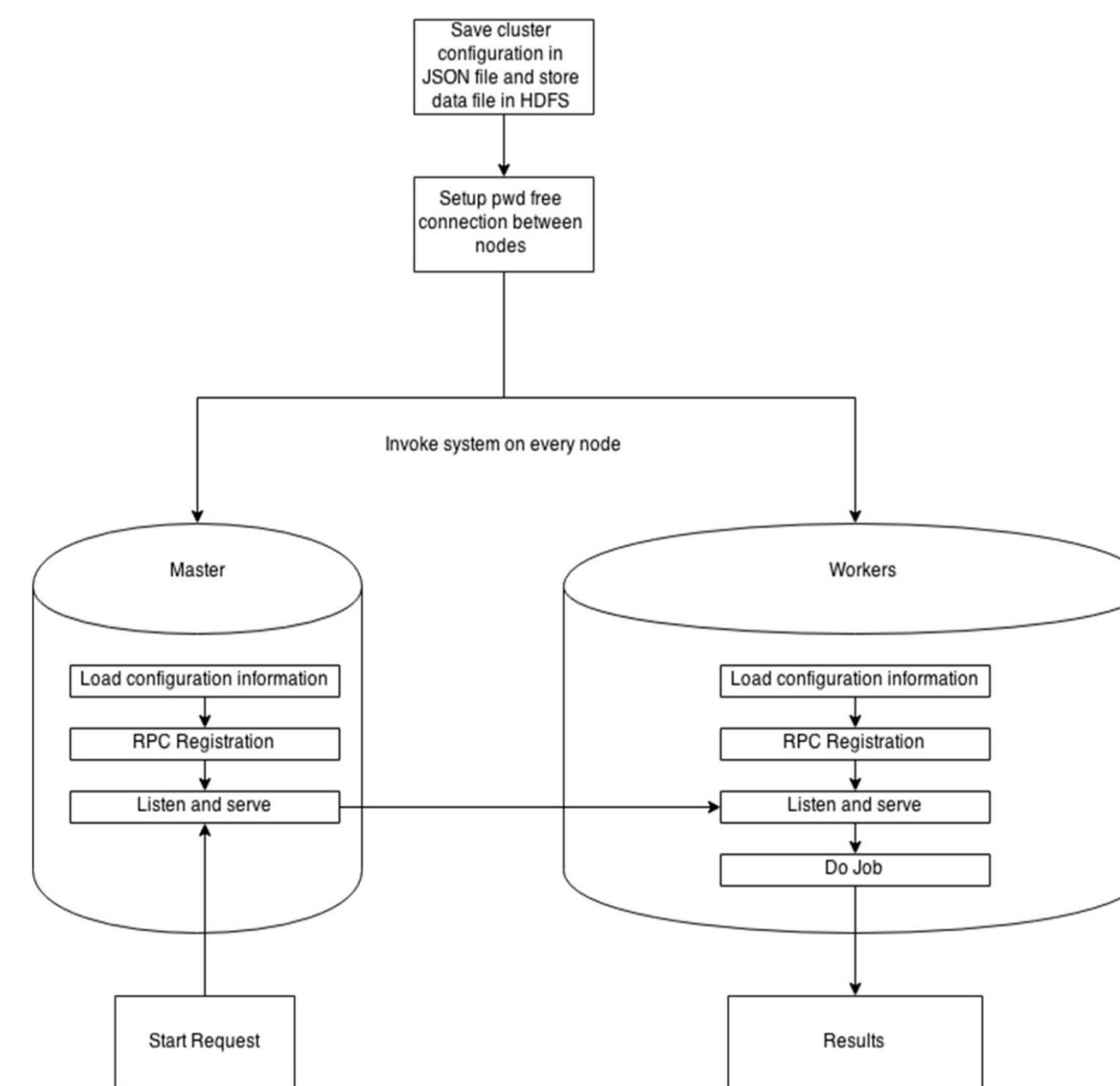
- We have designed and implemented a distributed computing system, which can manage a cluster of machines and run program in parallel among machines in the cluster. It is based on Remote Process Call (for program invocation) and Hadoop Distributed File System (for data share).
- Professor Long Quan's research group has developed a large-scale 3D Reconstruction program. Because of the huge dataset used, it usually has a significant large running time if run by one machine. This becomes the initiative of this project.



The system can:

- Load cluster configuration
- Generate input data and distribute them to individual machines.
- Command the workers to start execution of certain program.
- Check the running progress on all clients.
- Once finished, collecting all results.

Design and Implementation:

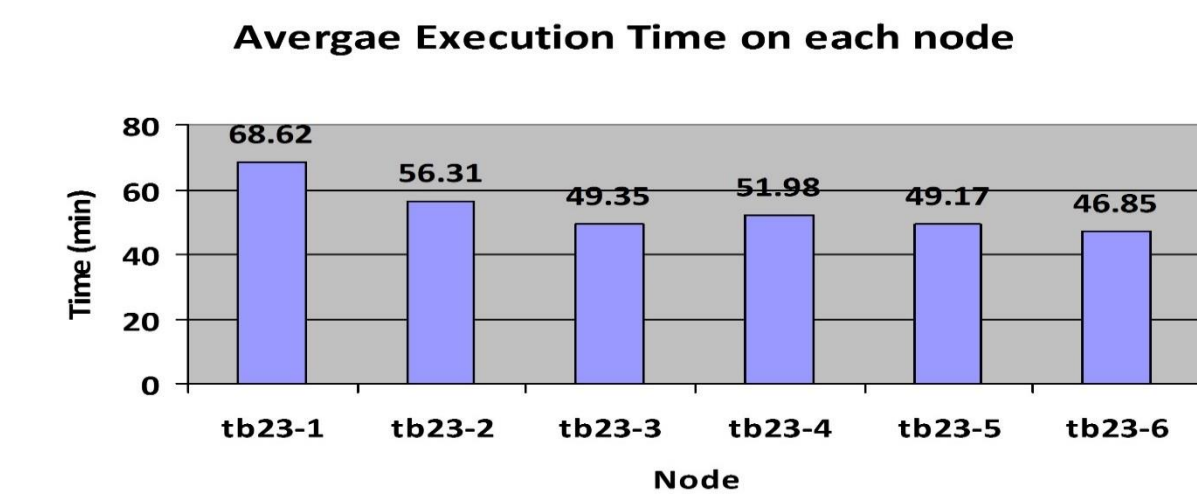


Workflow:

- Each node registers "DoJob" and "ListStatus" function
- Each node setups HTTP listener
- Master node received "start request" and dial remote RPCs
- Worker nodes start execution
- Results generated

Results:

- We first tested the execution time on each node. In the figure below, each column represents the average execution time on each node if only the node itself is used for execution (average taken over 5 executions).



- We then calculated the ideal time when running the program over a cluster of N nodes, simply by the formula below:

$$\frac{1}{N} \sum_{i=1}^N (\text{Average Execution Time on Node } n) / N$$

- After that, we recorded the time used to run program parallel on N nodes (N=2, 3, 4, 5, 6). Both the ideal time and actual time is shown in the graph below.

