# Scaling

**Name: Li Liu ID: LL2**

How to coordinate the roles

In my design, there is a master node, which is the first node started by system with id 1. And every time when another node is started, the master node will be notified and put the id of new node and its type in a map to record it. And when other nodes start, they will first ask master node what role they are and record that in their global value. By the way, the master node also serves as front end and cache VM.

How to start/shut down VM

First of all I will start another middle end VM and another front end VM at beginning. Before the first middle VM start working. all query come into master node will be dropped to avoid timeout error. And at that time, I will count how many requests are dropped to measure the rate of beginning, and start more middle VM according to that begging rate.

When all VM is running. I used some parameter to control start and shut down VMs. For front end, if there are too many requests in queue, which means front end is not enough, master node will detect that and start a new one. And for each front end, I will record the time it takes to get a requests, and shut down itself when this time is too long.

For middle end, if there are too many requests in central queue of master node, which will be check every time try to add request to that queue, the request will be dropped, and if it drops many time exceeding the threshold, master node will start a new middle end. What's more, every middle end will record the latest time they received a request, and check how long it passes in every loop, if it is long enough, the middle end will shut down itself.

Cache

I use master node as cache, and there is just a simple hash map. Every get operation will go to check hash map first. And every miss get operation or set operation will renew the map for further get. Other operations are just transferred by my cache.