Design Choices

1. The implementation of this program is fairly straightforward, it involves services for client <-> LB-server (client request), LB-server <-> server (memory messages) and server <-> client (direct messages) communications. All these servers are responsible for different tasks.

Running Details

1. The LB server needs to be run before the server or the clients. Following this, the user is prompted which policy should be implemented and this remains fixed throughout execution.
2. The client and server are then run on their respective terminals afterwards. The commandline arguments need to be correct.

Assumptions

1. The ETCD registry is happening at a fixed interval. This means that, as the question states, the servers send the LB server heartbeats at regular intervals. Hence, when a server is killed, and a client is immediately making a request, it is possible that the LB server assigns that client to a dead server. This is a rare but possible scenario, and it is handled by sending an error back to the client and assuming the client will eventually retry their request.
2. If the server fails in between the task going on, it will be okay to terminate all the processes that are running on that server.

Implementation Details

1. Round robin server maintains an internal index within the server and this is incremented (and modulus is taken) resulting in a cycling through. Pick first load balancing always chooses the first available server leading to possible underutilization of other servers. Least load policy utilizes `gopsutil` package to find out the memory on each port, and returns that memory consumed. The alive server with the least load is chosen ensuring a fair load balancing strategy.
2. As we were allowed to only use etcd for server discovery and sending heartbeat messages to keep the server alive, I am combining this with the memory sending service. Hence, the policies use an intersection of the memory information and the etcd server discovery, i.e. the least load policy will check only the memory usages of the alive servers. The other two policies do not require any memory information.