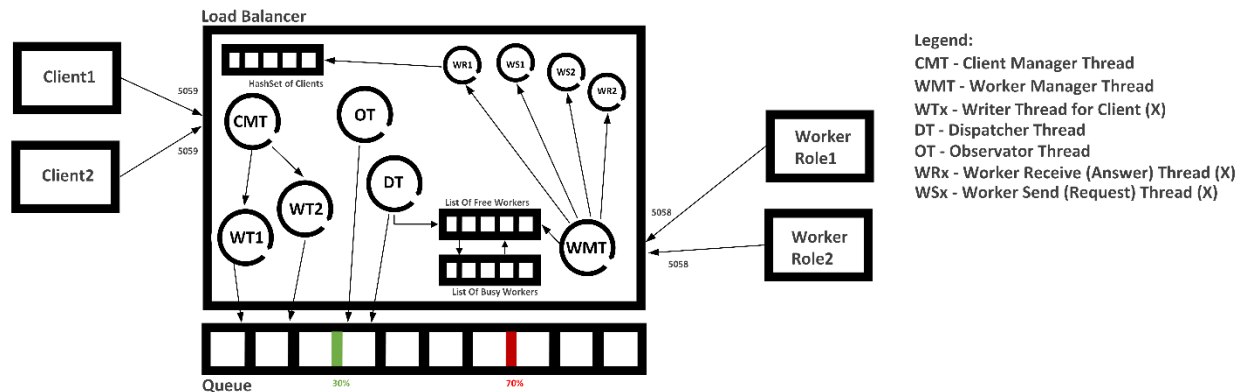


IKP – Dizajn projekta za balansiranje opterećenja



Client se prijavljuje na aplikaciju preko porta 5059 i šalje zahteve za obradu. **Client Manager Thread** prijavljenog korisnika smešta u HashSet za klijente, pravi novi **writer thread** i dodeljuje ga korisniku. Preko ovog thread-a zahtev korisnika se smešta u **Queue** (kružni buffer).

Observer thread proverava popunjenost queue-a u određenom vremenskom intervalu i na osnovu toga kontroliše broj instanci worker rola na sledeći način:

1. Ukoliko je popunjenost preko 70% - OT pravi novu worker rolu da bi smanjio opterećenje.
2. Ukoliko je popunjenost manja od 30% - OT isključuje jednu od slobodnih instanci dok broj instanci ne bude 1.
3. Ukoliko je popunjenost između 30% i 70% neće biti promena.

Kreirana worker rola se prijavljuje na aplikaciju na port 5058. Tada, **Worker Manager Thread** dodeljuje mu dva threada – **send thread** i **receive thread** i informacije o worker roli (o thread-ovima) smešta u **listu slobodnih workera**.

Dispatcher thread čita zahteve iz queue-a i dodeljuje zahtev prvoj instanci workera iz **liste slobodnih workera**. Za odgovarajućeg workera **send thread** šalje zahtev instanci workera za obradu i tada se prebacuje u **listu zauzetih workera**. Kada instanca workera završi obradu zahteva, njegov **receive thread** prima odgovor i prosleđuje ga odgovarajućem klijentu u HashSet-u. Nakon toga, instanca workera se vraća u **listu slobodnih workera**.

Client će biti aktivan na aplikaciji dok god ne pošalje komandu "quit".