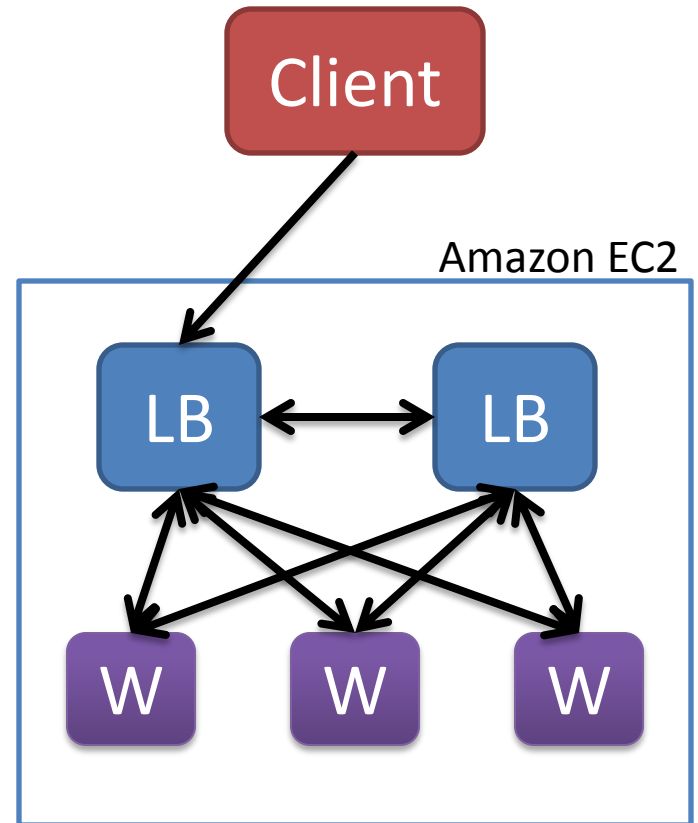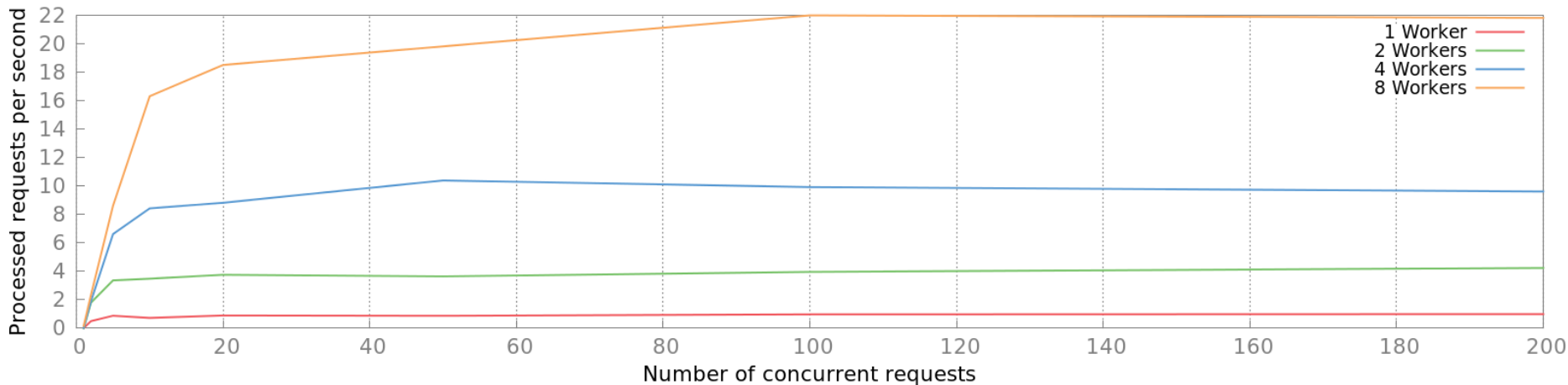# Context

- Requests per second:
  - # of Concurrent requests
  - # of servers
- Response time
  - Autoscaling
  - # of servers

Benchmark

- Client on US West N. California

- System on US West Oregon
  - Connection latency
  - Low overhead

Autoscaling