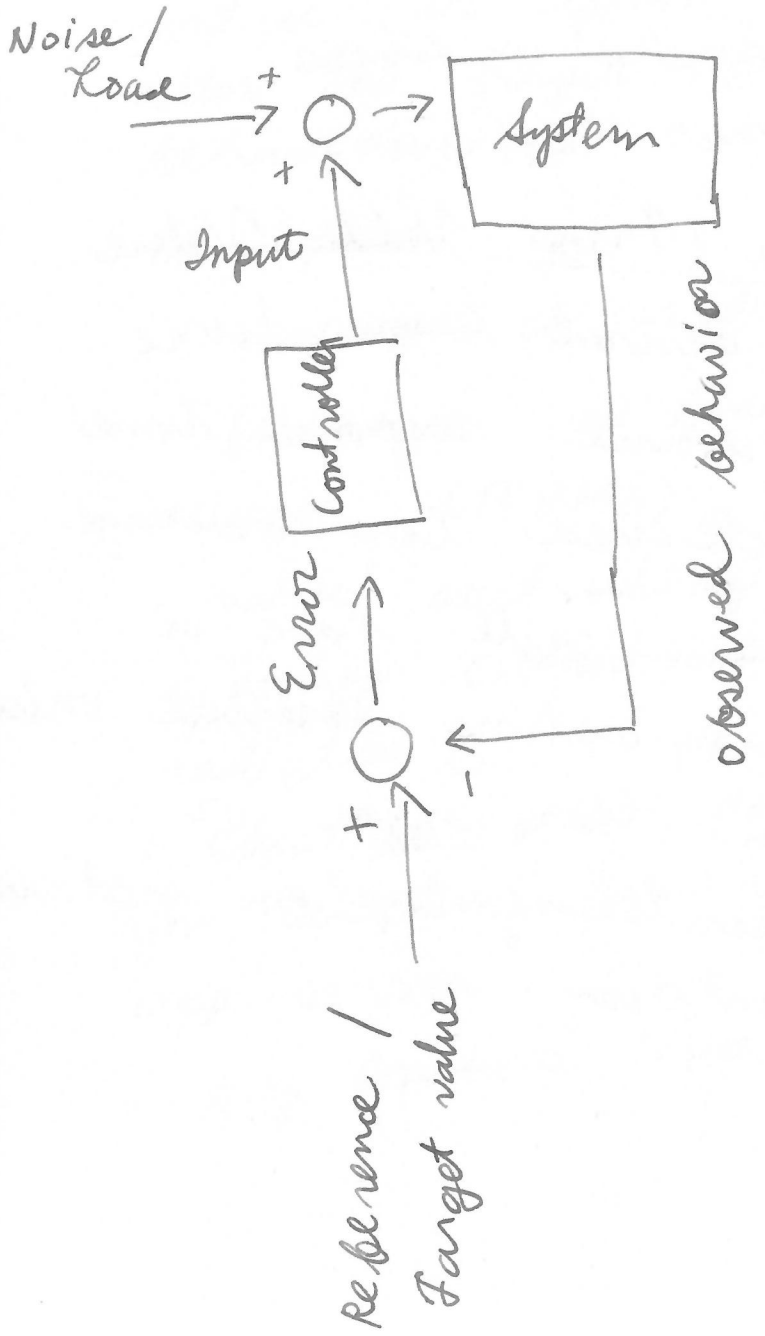
In this talk, I will show how we can use the same concepts to manage and autoscale the health of our applications on Kubernetes.

Feedback Architecture:



Noise / Load:

- weather conditions
- web traffic

Reference / Target ~~Value~~: Value:

- desired temperature
- desired response time
- desired load average

→ (each will have a corresponding observed value from the sensors or monitoring system instrumentation.

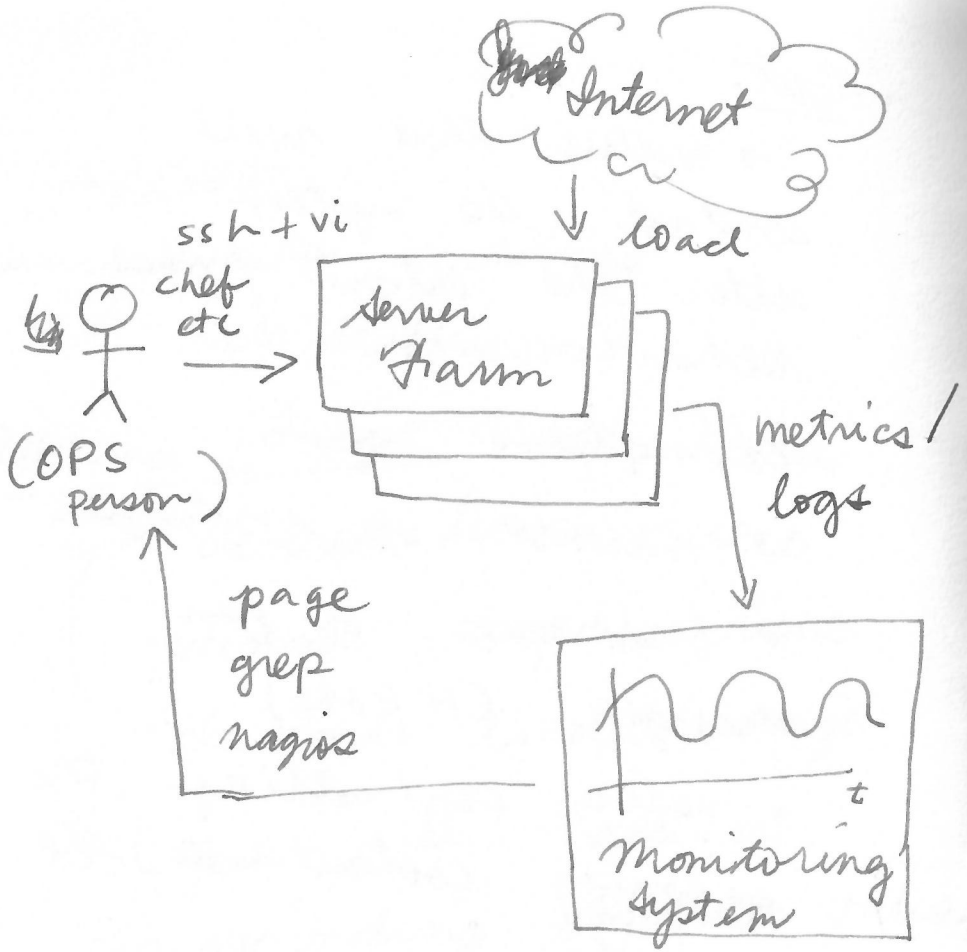
Input:

→ value that will control the system to alter the output / observed behaviour

Examples:

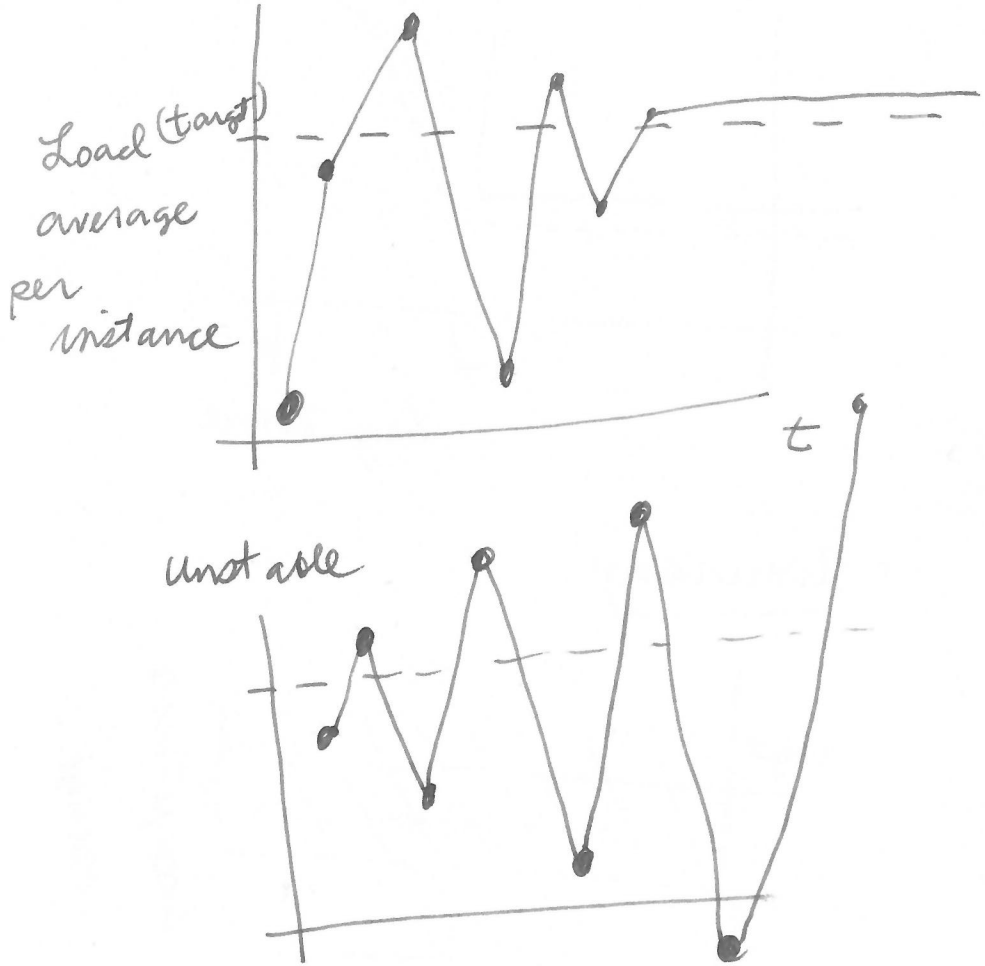
- number of instances
- furnance or fan strength (RPMs)
- valve controlling the incoming coolant to the radiator.
- can be other parameters you can tune for your app server

  e.g. Apache's MaxClients.

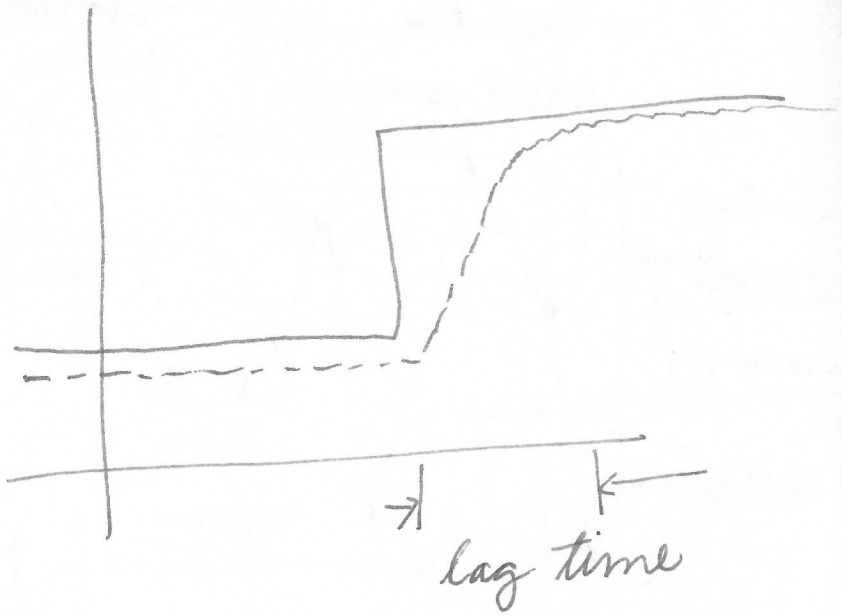


Error correction

- Stability
  - → results from over-correction



- Performance
  - response to change



- accuracy