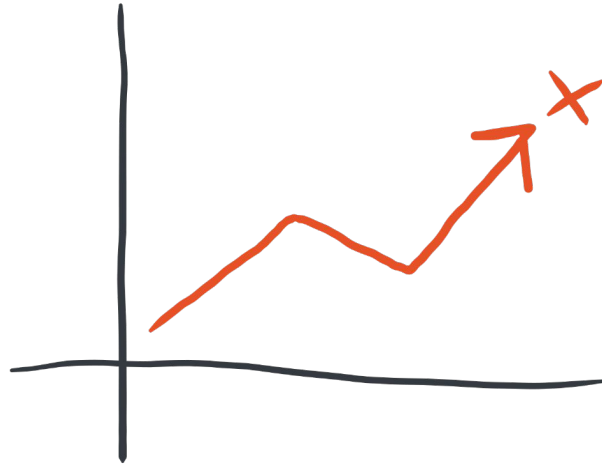


Service Level Objectives

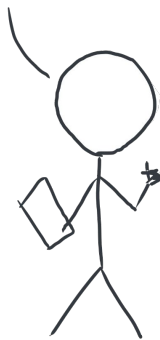


It's the **user experience** that matters

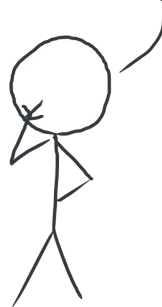


Define what you **really** want to guarantee

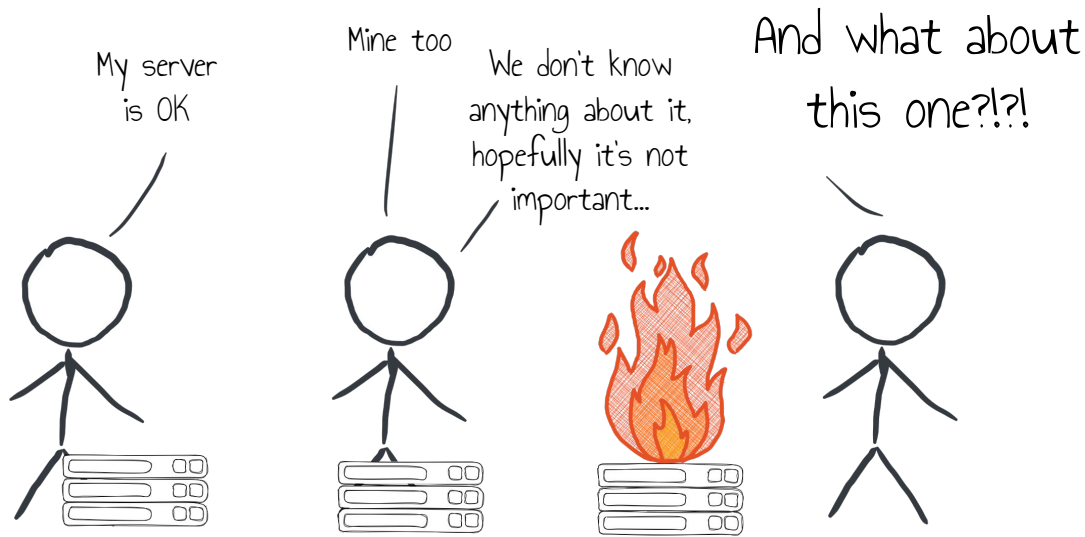
We need 100% uptime 24/7 over 365 days a year, any operation cannot take longer than 15 ms and there must be someone on call all the time!



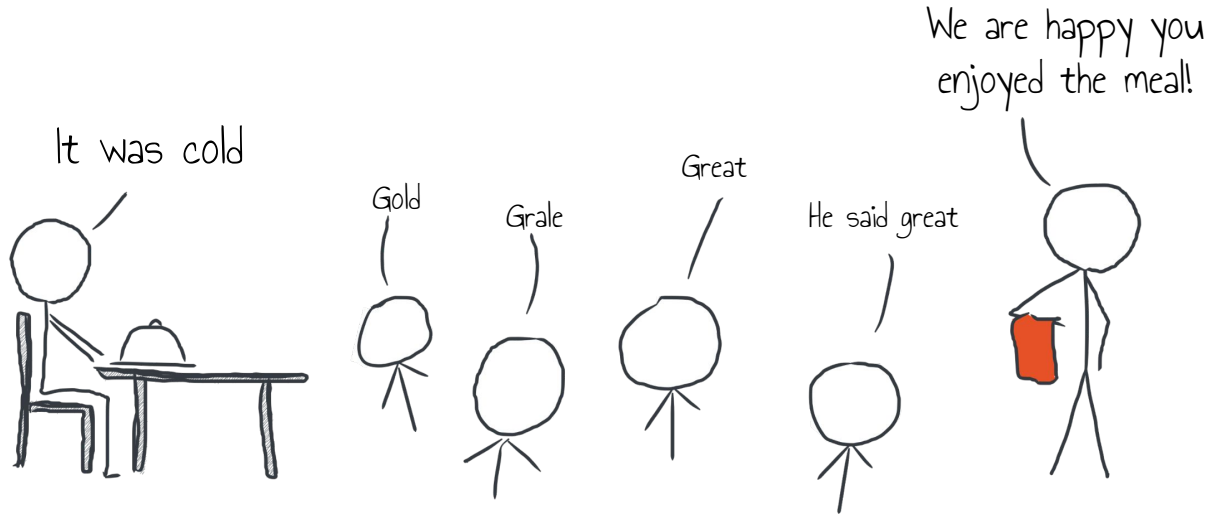
Really? You know the service just generates greeting cards for a Mothers day once a year right?



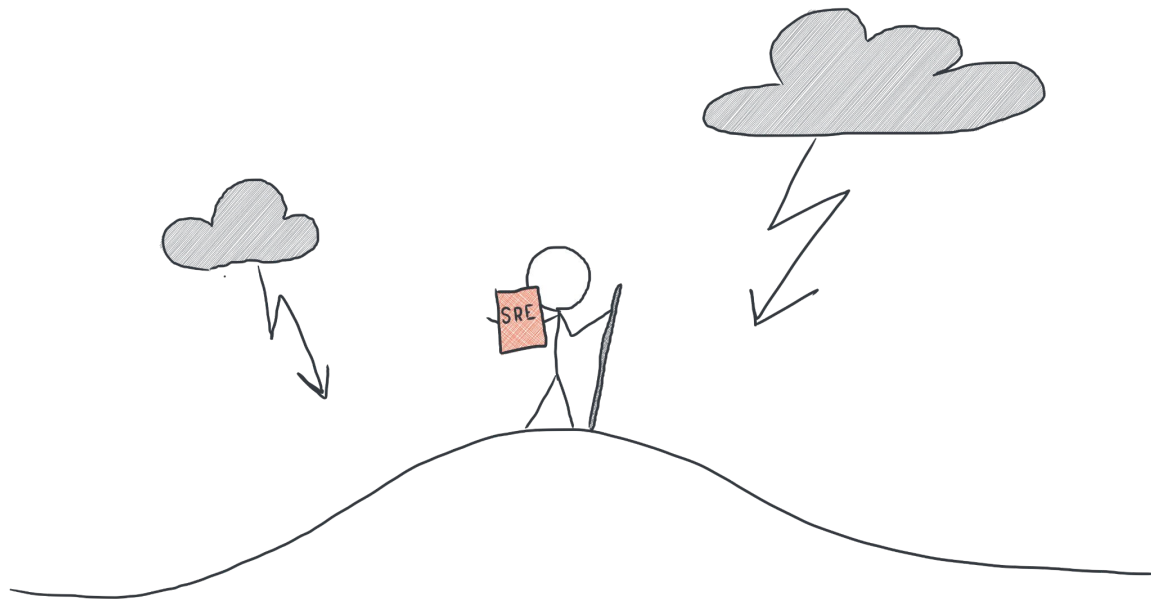
Observe the service as a whole



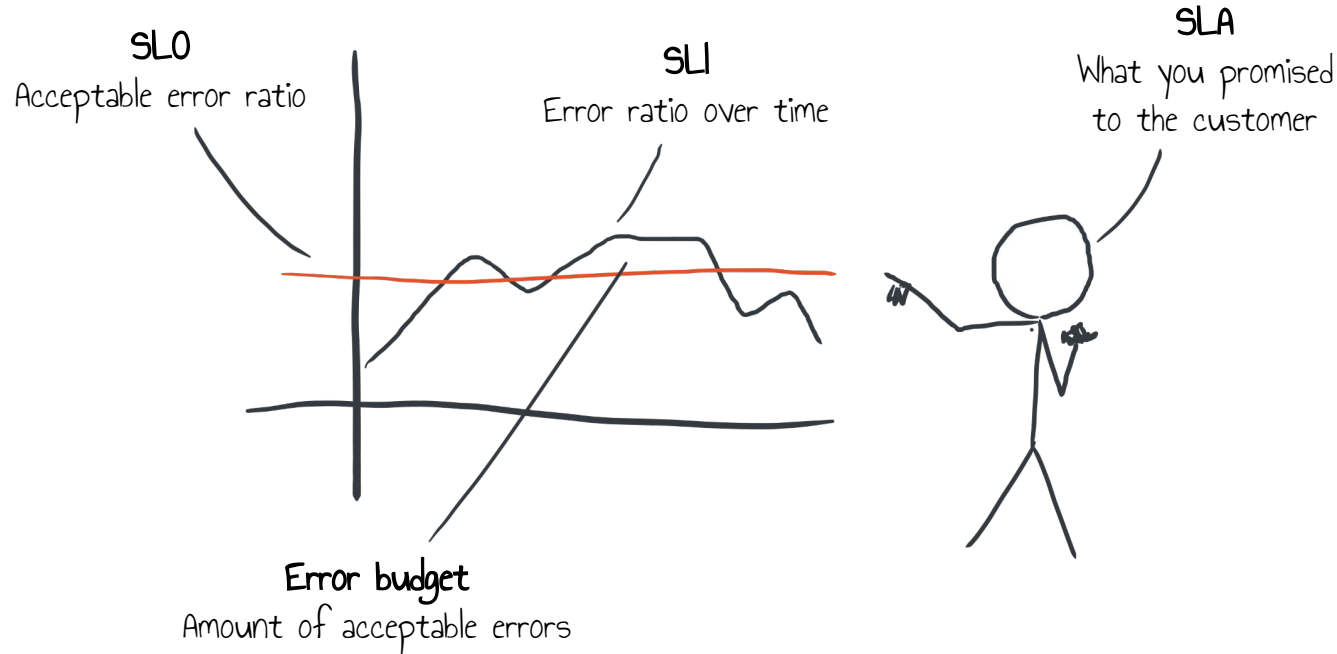
Evaluate impact on the user
as close as possible



SRE books from Google



Service Level Objectives terminology



Implementation using Prometheus

slo:error_ratio1h

$$\frac{\text{sum}(\text{rate}(\text{http_requests_total}\{\text{app}=\text{"foo"}, \text{status!}\sim\text{"5.."}\}[1\text{h}]))}{\text{sum}(\text{rate}(\text{http_requests_total}\{\text{app}=\text{"foo"}\}[1\text{h}]))}$$

Implementation using Prometheus

Multi-window burn rate alert

`slo:error_ratio1h{app="foo"} > (14.4*0.001)`

and

`slo:error_ratio5m{job="myjob"} > (14.4*0.001)`

or

`slo:error_ratio6h{job="myjob"} > (6*0.001)`

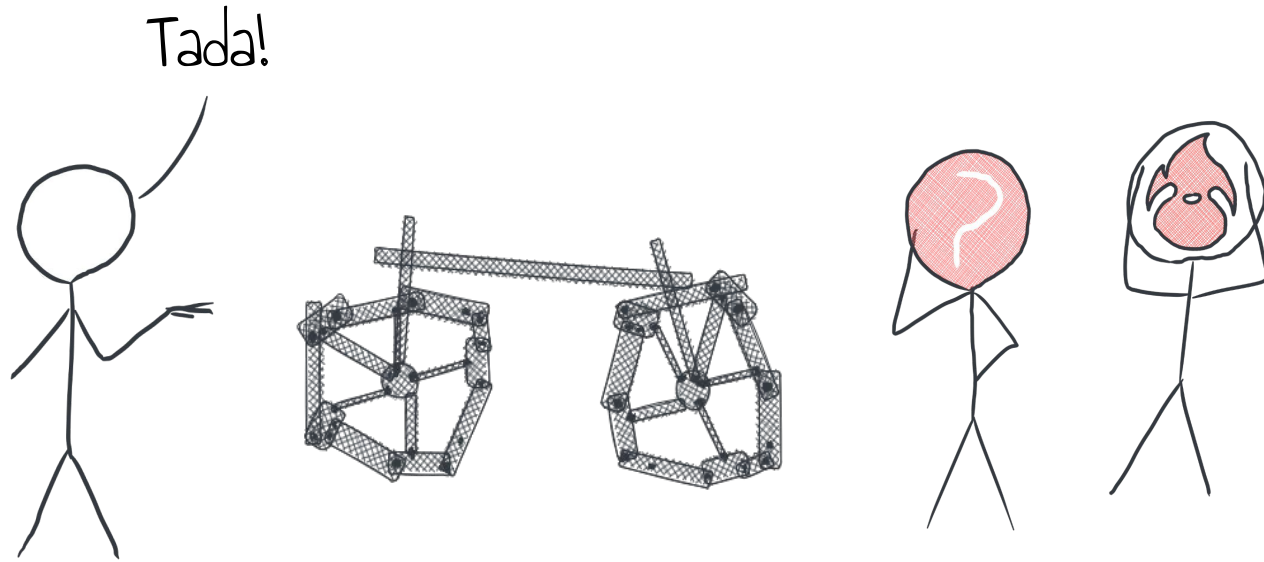
and

`slo:error_ratio30m{job="myjob"} > (6*0.001)`

Burn rate is
how fast you are burning the
error budget



Do not reinvent the wheel





Links

[Pyrra](#)

[Sloth](#)

[slo-libsonnet](#)

[slo-exporter](#)

[OpenSLO](#)