Four Golden Signals

Monitoring the health of your service
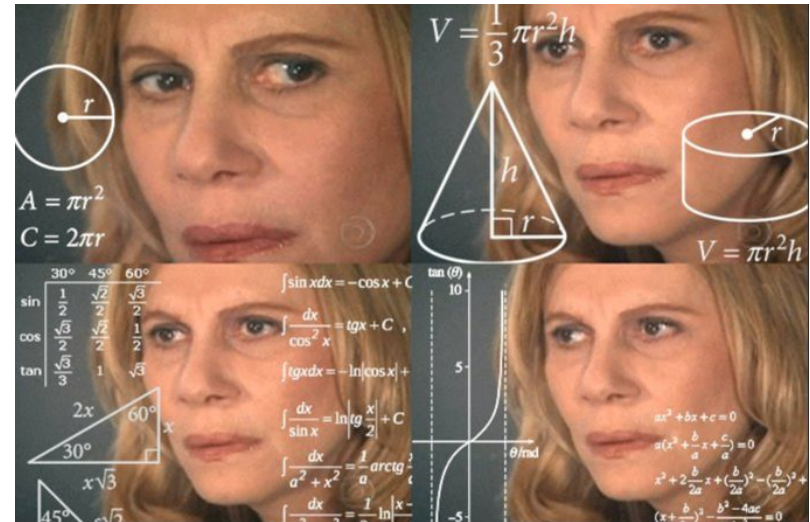
Why Do We Monitor?

- Understand long-term trends

- Understand the impact between changes

- Alerting

- Debugging & Troubleshooting

Making sense of it all

- Host level metrics

- Infrastructure metrics (Load Balancer)

- Application level metrics

Each level could have 100s (or more!) different metrics you can monitor!
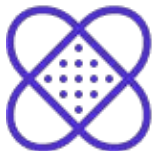
The Four Golden Signals

Latency

Traffic

Errors

Saturation

Latency

How long does it take to service a request?

Not all requests are equal, so it may not be appropriate to treat them as such

Traffic

A system level metric indicating how much demand is being placed across your service.

For a web service, this would be HTTP requests per second.

Errors

The rate of requests that fail, this may be because of an explicit error (like a HTTP 500) or perhaps implicitly, such as serving the wrong content

Saturation

How close to the limit your service is in terms
of resource constraints, such as memory, CPU,
or I/O

Putting It Into Practice