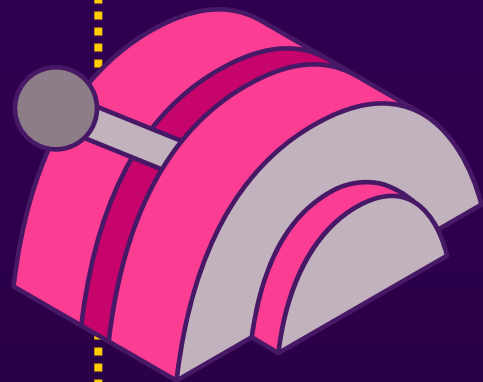


# SRE Availability Metrics



## 9's of availability



Uptime	Downtime (per year)
99%	3d 15h 39m
99.9%	8h 45m 56s
99.99%	52m 35s
99.999%	5m 15s
99.9999%	31s
99.99999%	3s

## Jargon

Industry jargon refers to the number of "9's" related to availability.

For instance, one 9 would be 90%, while five 9's would be 99.999%



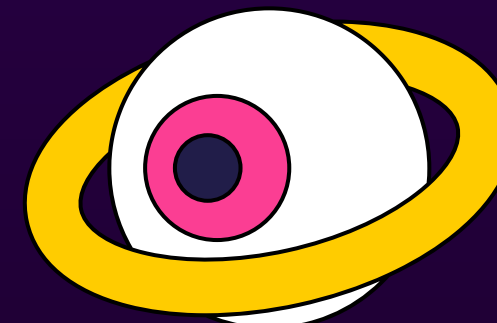
## Service Monitoring

For services-related monitoring consider the **RED** method.

**Rate** (the number of requests per second)

**Errors** (the number of those requests that are failing)

**Duration** (the amount of time those requests take)



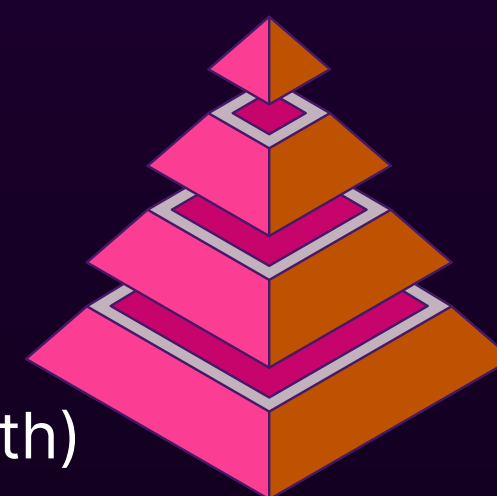
## Hardware Monitoring

For hardware-related monitoring consider the **USE** method.

**Utilization** (% time that resource was busy)

**Saturation** (amount of work resource has to do, often queue length)

**Errors** (count of error events)



## Kubernetes Monitoring

For Kubernetes-related monitoring consider the Four Golden Signs

**Latency** (time taken to serve a request)

**Traffic** (how much demand is placed on your system)

**Errors** (rate of requests that are failing)

**Saturation** (how "full" your service is)

