Example SLO Document

This document describes the SLOs for Cat Facts service.

Status: Published

Date: 6/5/2017

Reviewers: Marcus Rogers

Approvers: Amanda Bleck

Approval Date: 6/7/2017

Revisit Date: 6/1/2018

Service Overview

The Cat Facts service allows Android and iPhone users to retrieve facts about cats. The app runs on users' phones and new facts are retrieved from the API via RESTful, HTTP-based requests. The data store contains the cat facts along with relevant metadata. Cat facts are available in the app, via the API, and also on a public HTTP server.

The SLO uses a four-week rolling window.

SLIs and SLOs

Category	SLI	SLO
API		
Availability	Proportion of successful requests measured via load balancer metrics. All HTTP status codes except 500–599 count as success. count of "api" http_requests NOT 5xx status code / count of all "api" http_requests	98% success
Latency	Proportion of sufficiently fast requests measured via load balancer metrics. Sufficiently fast defined as < 500 ms. count of "api" http_requests where duration < 500 ms / count of all "api" http_requests	95% of requests < 500 ms

HTTP Server		
Availability	Proportion of successful requests measured via load balancer metrics. All HTTP status codes except 500–599 count as success. count of “web” http_requests NOT 5xx status code / count of all “web” http_requests	95% success
Latency	Proportion of sufficiently fast requests measured via load balancer metrics. Sufficiently fast defined as < 500 ms. count of “api” http_requests where duration < 500 ms / count of all “api” http_requests	95% of requests < 500 ms

Rationale

SLIs were based on measurements over the period 5/1/2017 to 5/31/2017. Availability SLOs were rounded down to the nearest 1% and latency SLO timings were rounded up to the nearest 50 ms.

Error Budget

Each objective has its own separate error budget, defined as 100% minus the goal for that objective. For example, if there have been 1,000,000 requests to the HTTP server in the previous four weeks, the HTTP availability error budget is 5% (100%–95%) of 1,000,000: 50,000 errors.

Additional Information

- Request metrics are measured at the load balancer. Because of this, the metric may fail to accurately measure requests that do not make it to the load balancer.
- Only 5xx status codes are counted as errors, all other requests are counted as successes.