Defining Smart Performance Objectives for APIs

Nick Denny
VP Engineering & Co-founder

@nickdenny
nick@apimetrics.com
“And what is good, Phaedrus, And what is not good— Need we ask anyone to tell us these things?”
What is “good” when we talk about APIs?

- SLAs
- SLOs
- KPIs

& how to measure them
Some definitions

**Service Level Agreements (SLAs)**

- Goals set to manage performance between different parties i.e. the API provider and the user
Some definitions

Service Level Objectives (SLOs)

• Technical measures you set as goals
Some definitions

Key Performance Indicators (KPIs)

• Items you track to understand how you are doing and where to improve
So...

- **An SLA is** we will meet SLO

- **The SLO is** Five 9s availability on a weekly basis

- **Our KPIs are** how does our availability this week compare to last week and where did we miss?
Who are you measuring for?

- For customers?
- For management?
- For regulators?
- Because somebody told you to measure something?

**Monitoring tells you what you are doing and how you can improve it...**

**don’t confuse it with testing...**
KPIs – Hard and Soft

- Key Performance Indicators aren’t always technical
- Soft KPIs include:
  - Adoption Rates
  - How much are the APIs being used?
  - Are they making money?
- John Musser gives a great talk on this... worth finding on YouTube!
- We’ll stick to the technical ones
What you can measure with an API

- HTTP Code
  - The crudest measure
- Latency
  - Also a bit crude
- Geographic factors
  - Getting warmer
- Content validation
  - Much better
- Overall performance trends & blended metrics
  - Better still
HTTP Codes

HTTP/1.1 200 OK

(Except when it’s not)
Latency

What are you measuring?
And where are you measuring it from?

API Call Breakdown (Median Call Length)

- **External**
  - DocuSign
  - Box

- **Internal**

Despite being labeled as 'API Call Breakdown', the graph simplifies the process into a single bar chart. It appears to illustrate the breakdown of API calls into various components such as DNS, Connection, Handshake, Upload, Processing, and Download. The components are color-coded for easy differentiation.

The bar chart suggests that the process is more complex than initially assumed, as indicated by the label 'Look, it’s complicated'.
Geographic & Cloud Factors

Global Cloud Performance

- AWS: 532
- Google: 537
- Azure: 548
- IBM: 499

North America

Europe

Asia

SE Asia

South America

Oceania
Latency isn’t everything...
Content Validation
Trend Based

Bank Overall API Quality Score May-July 2018

Allied Irish
Bank of Ireland
Bank of Scotland
Barclays
Danske
Halifax
HSBC
Lloyds
Nationwide
NatWest
RBS
Santander
Ulster
Service Availability & SLOs

- How long actually is an outage?
- If your sample is too small, can you miss things?
- Gateway or Server logs miss external factors & load issues

Need to define service availability SLOs careful to take into account network issues and the method of measuring outage
What makes an SLO smart?

- "You can only manage something if you can measure it"
- It's easy to obtain a lot of data
  - Server, Gateway and other logs
- But an SLO needs to tell all interested parties what they need to know
- Many factors affect API performance
  - Performance issues might be out of your control
    - But? How do you prove that to a paying customer... like... randomly, CISCO?
    - Or a regulator like the OBIE in the UK?
Anomaly detection and SLOs

- Goal: API to behave within specification and have no (or minimal) anomalies
- Apply **heuristics** or using **machine learning** to characterize API behavior (your trends) and detect anomalies (outliers)
- Understanding API better allows more intelligent and smarter SLOs to be set
- Smarter SLOs make for achievable SLAs
Things to look at when measuring

- Latency from where a call likely originates
- Location & Cloud
  - Is the problem in the cloud?
    - Bad DNS settings
    - Your CDN is lying to you
    - Some other ‘weird’ factor
  - Is the problem in your servers
  - Is it a weird combination of the above?
    - i.e. your servers barf on calls made from Azure Korea and ONLY Azure Korea?
Detecting latency component outliers

Distribution v Percentiles
Detecting latency component outliers

Distribution v Percentiles
**Smart** and **Intelligent** SLOs

- An SLO is **smart** if it relates to some business/operational issue that matters to the API provider and/or users.
- An SLO is **intelligent** if some intelligence (human or artificial) is applied to setting the SLO and then determining whether it has been met.
- **Don't** set an SLO just because a metric can be easily measured.
- **Don't** set SLOs just because you know they will be met (or not met).
- **Do** think **smartly** and **intelligently** about what SLOs to set, how to measure them and how they can be used to improve API quality continuously (i.e. watch the KPIs you set).
A quick plug (sorry!)

APImetrics does all this in a single product...

Key thing... **use something** – it doesn’t have to be us  ...but use something to track this – **don’t assume** some other part of your stack monitoring is doing it
Answers, Not Excuses

Engage with us today
Nick Denny
@nickdenny
nick@apimetrics.com