

Achieving **Excellence:** Thresholds that Transform Service Quality

Thiara Ortiz
Staff CDN Reliability Engineer
October 29, 2024



Episode 1

What's an SLO?

SLI & SLO

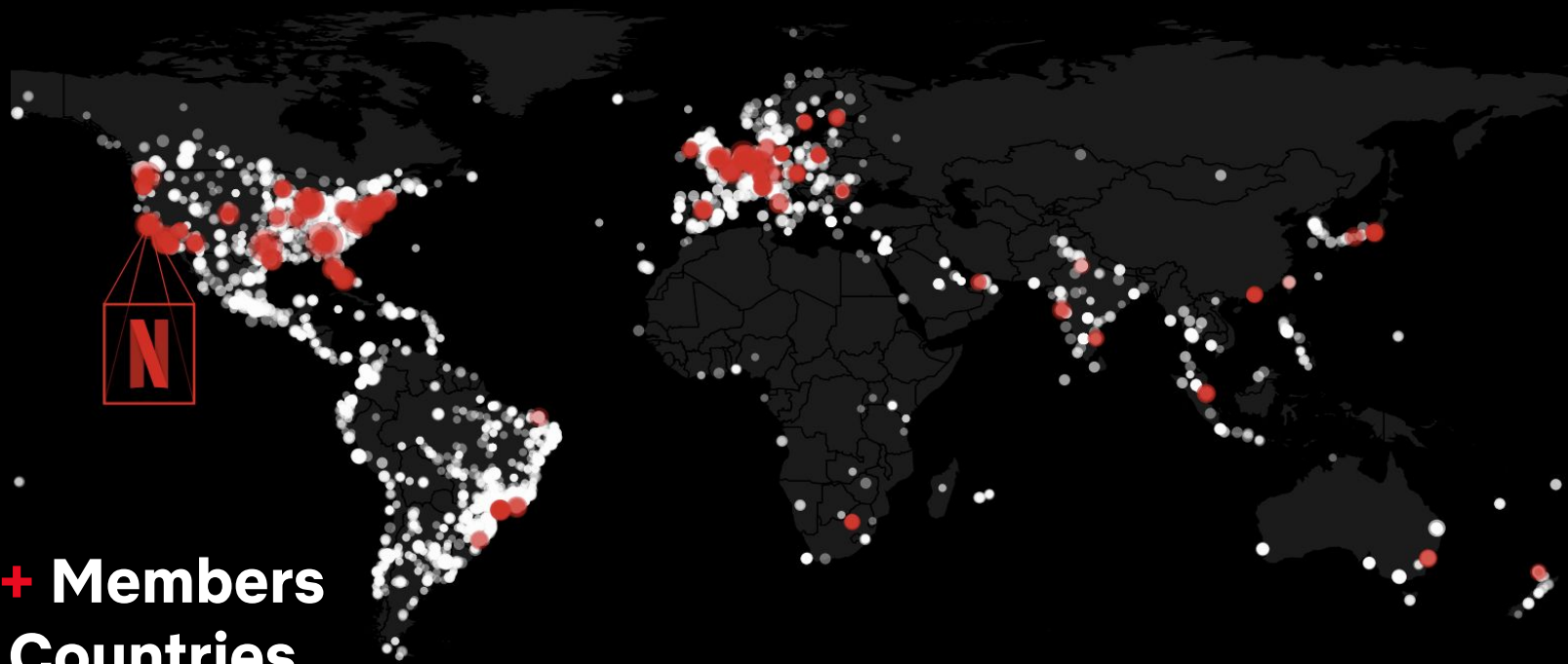
Service Level Indicator (SLI): quantitative measure of an aspect of a service

Ex: Request Latency

Service Level Objective (SLO): A target range of values for a service

Ex: Request Latency < 50 ms

Open Connect

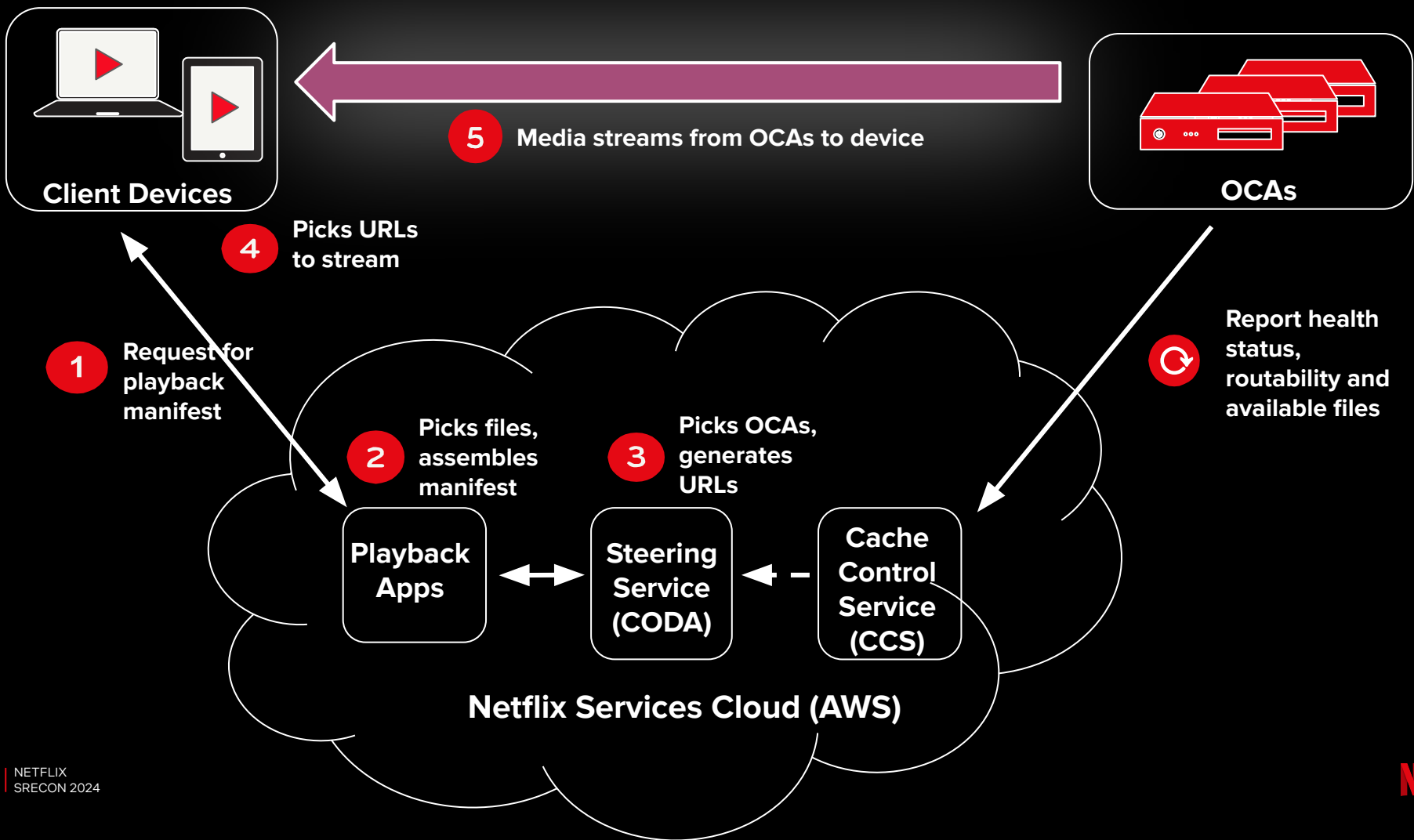


230M + Members

175+ Countries

18K+ OCAs

N

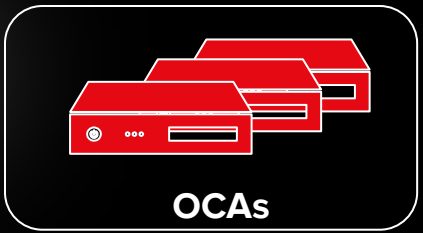




Client Devices



5 Media streams from OCAs to device



OCAs

Quality of Experience

Picks URLs
Time to URLs

Duration < X ms

manifest

2 Picks files, assembles manifest

3 Picks OCAs, generates URLs



Netflix Services Cloud (AWS)



Report health status, routability and available files

Error Rates

% errors < .05%

Episode 2

Quality of Experience

QoE

A measure of satisfaction with a telecommunication service. It focuses on the whole experience for a member.

At Netflix, these are metrics that we would monitor that encompass the user experience during playback.

NETFLIX





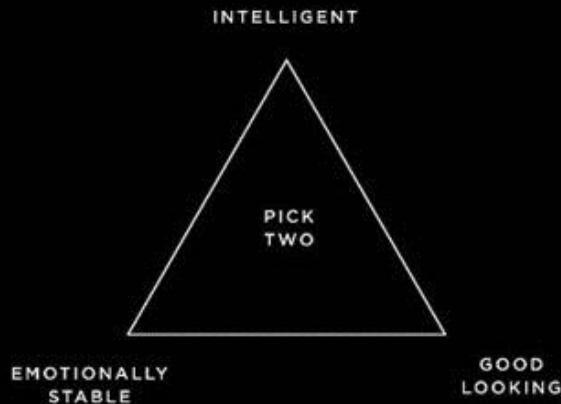
NETFLIX

▶ ⏮ 🔊 0:00 / 1:22 🔍 ⚙️ 📺 🖥️ 🗑️

QoE Tradeoffs

We collect logs to be able to quantify the following for each streamed session:

1. Play Delay
2. Network Rebuffers
3. Reduced Bitrate



Episode 3

Defining Thresholds for QoE SLOs

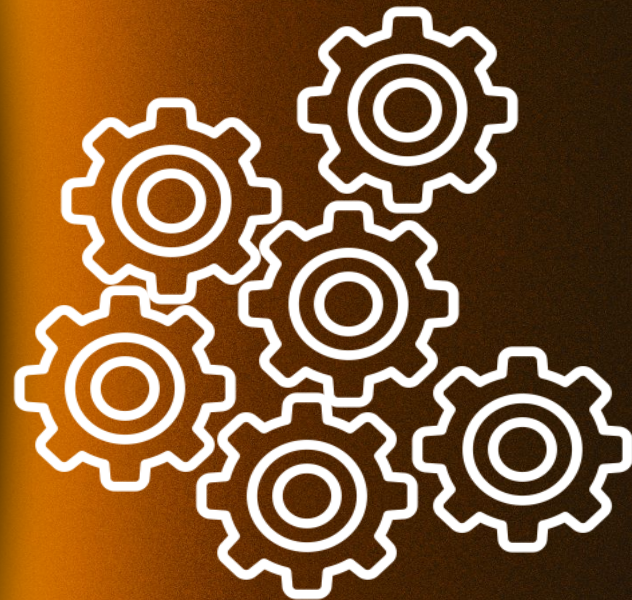
Level 1



Level 2



Level 3



Level 1



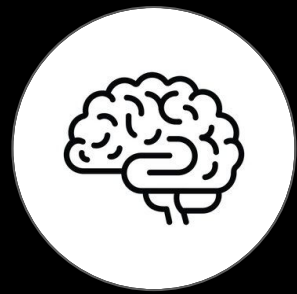
Level 2



Level 3



Level 1: **Intuition**



*What would I deem as **acceptable** as a user of the service?*

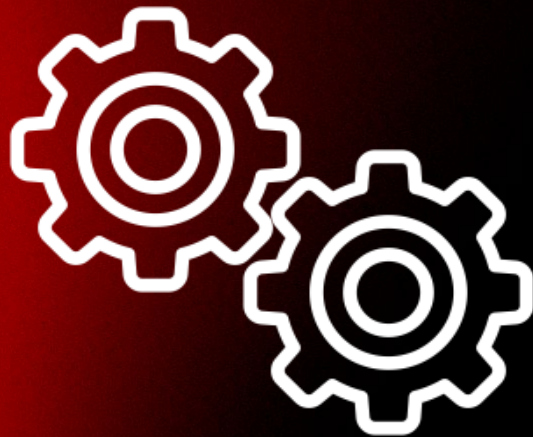
Pro: Easy

Con: May not be a good threshold for sensitive applications

Level 1



Level 2

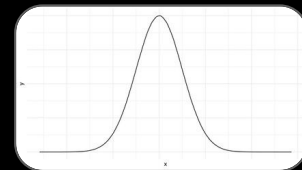


Level 3



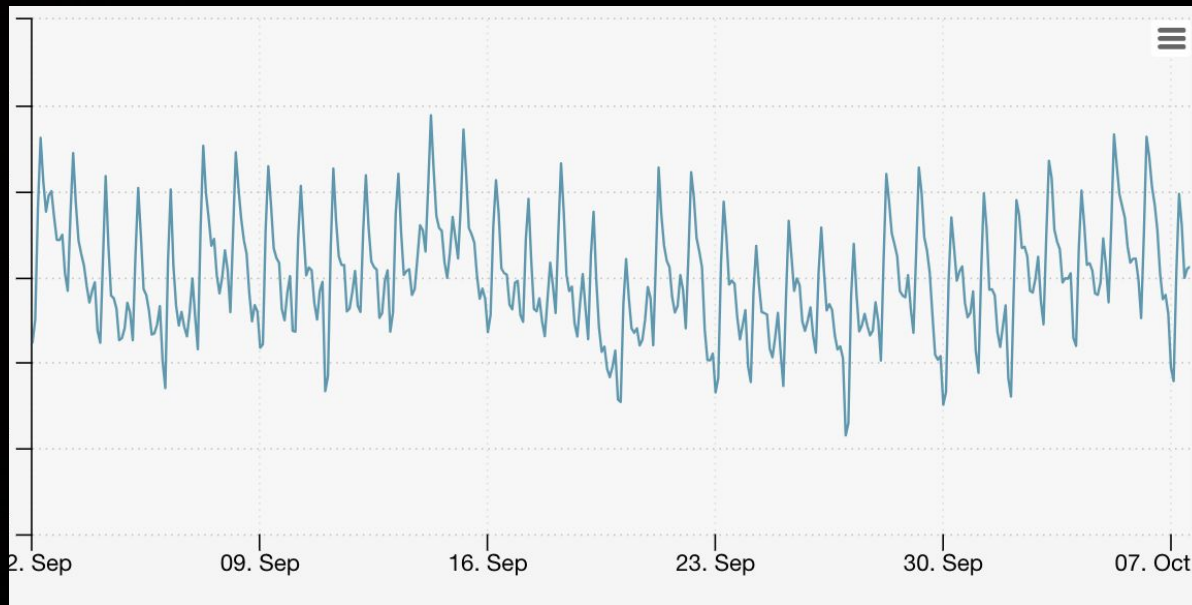
complexity

Level 2: **Historical Data**

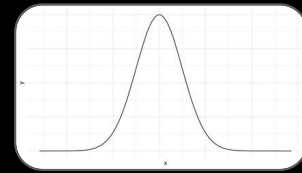


What does the past tell me about the capabilities of my service?

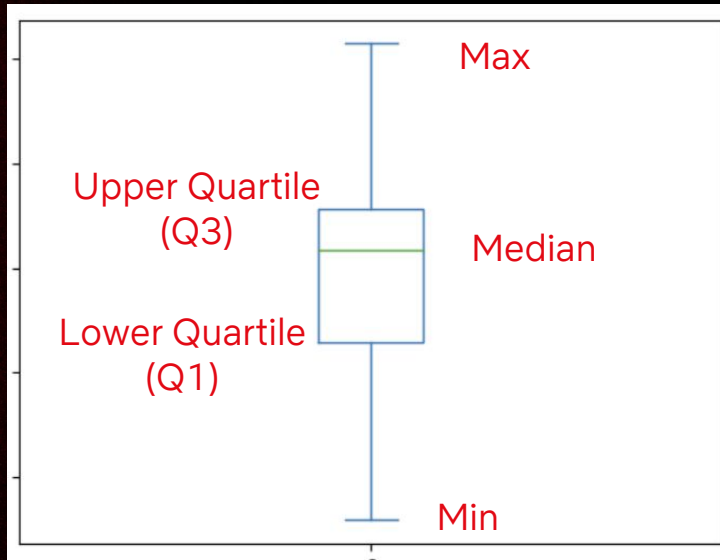
Normalized Bitrate



Level 2: Historical Data



Learn from the distribution of the data:

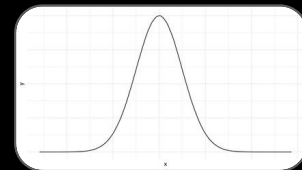


Outlier is a data point that is significantly different from the other values in the dataset

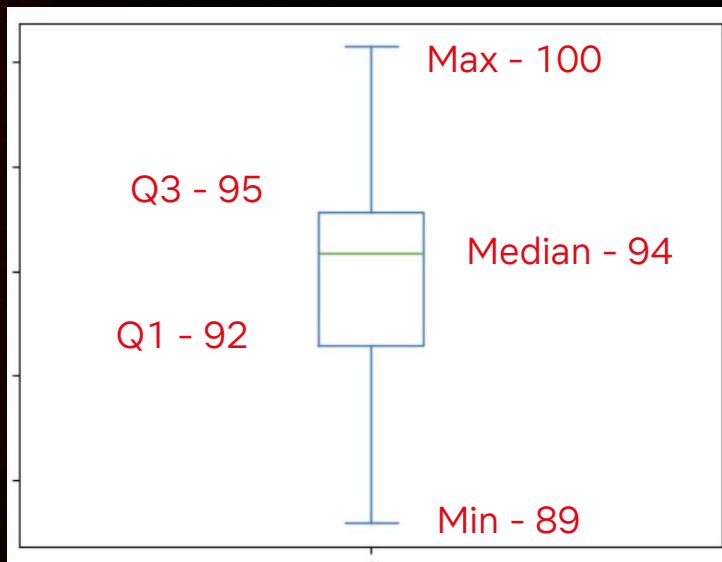
Interquartile Range = $Q3 - Q1$

$Q3 + 1.5 * IQR$ **or** $Q1 - 1.5 * IQR$

Level 2: Historical Data



Hypothetical Normalized Bitrate



$$\text{Potential SLO} = Q1 - 1.5 * IQR$$

$$IQR = 3$$

$$92 - 1.5 * 3 = 87.5$$

$$\text{Potential SLO} \leq 87.5$$

Level 2: Historical Data



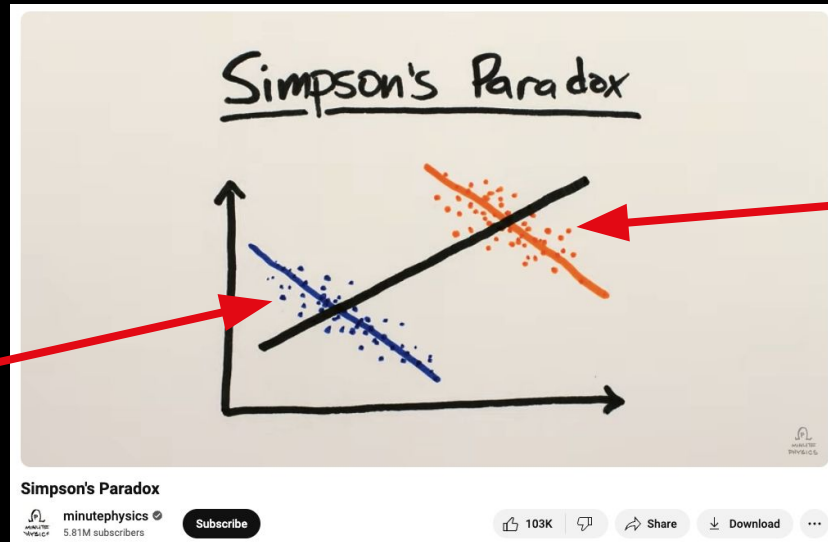
Warning: Consider the factors that make up the aggregate view



Differences in Device Mixture

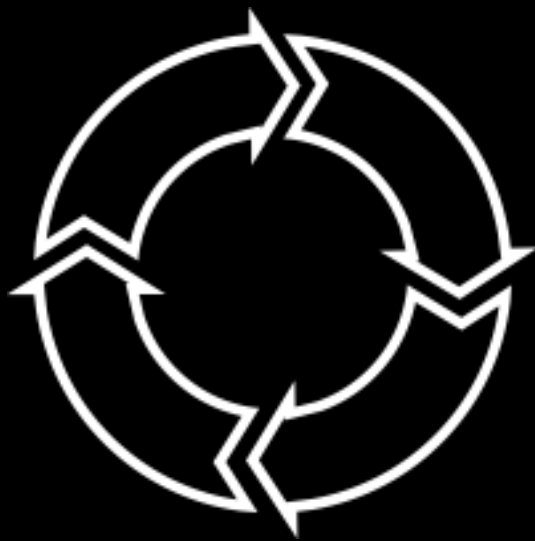
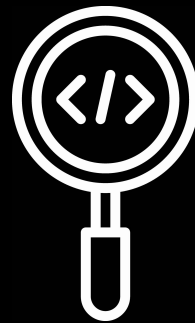
Simpson's Paradox

A phenomenon in statistics in which a trend appears in several groups but disappears or reverses when the groups are combined.



Level 2: **Historical Data**

Sanity Checking Against Incidents



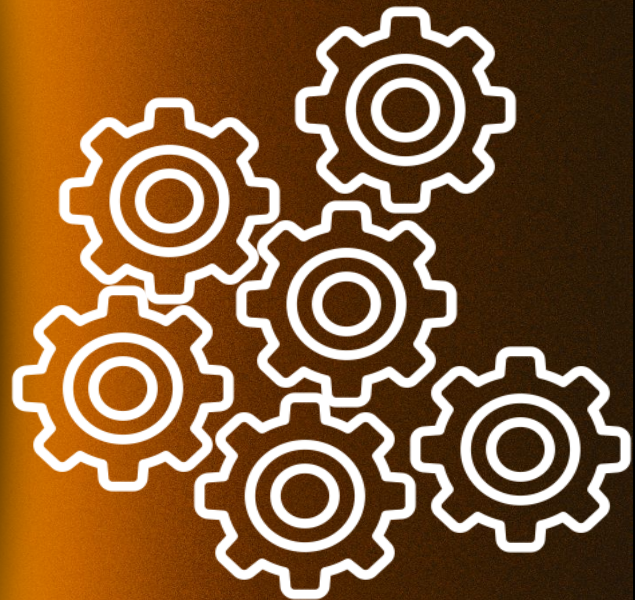
Level 1



Level 2



Level 3



Level 3: Implementing Surveys



Scenario: You're watching Netflix :-)

Demo: Everyone Raise Your Hand

Lower your hand if a single rebuffer would be enough for you to stop watching

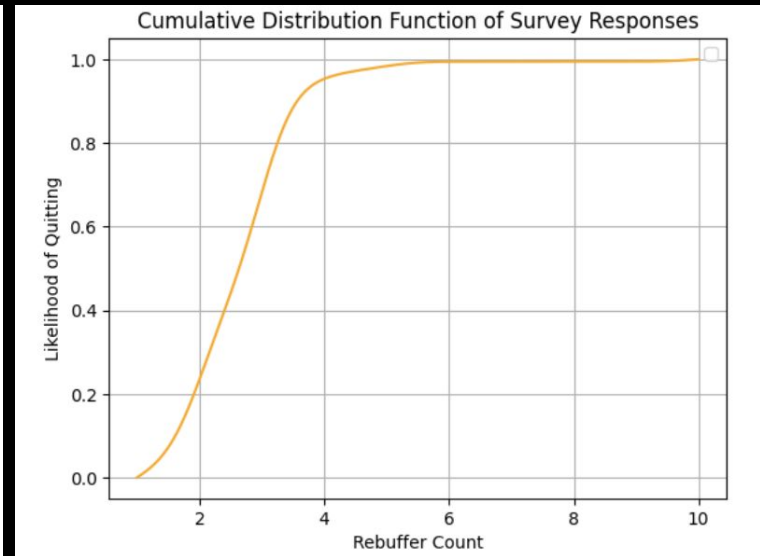
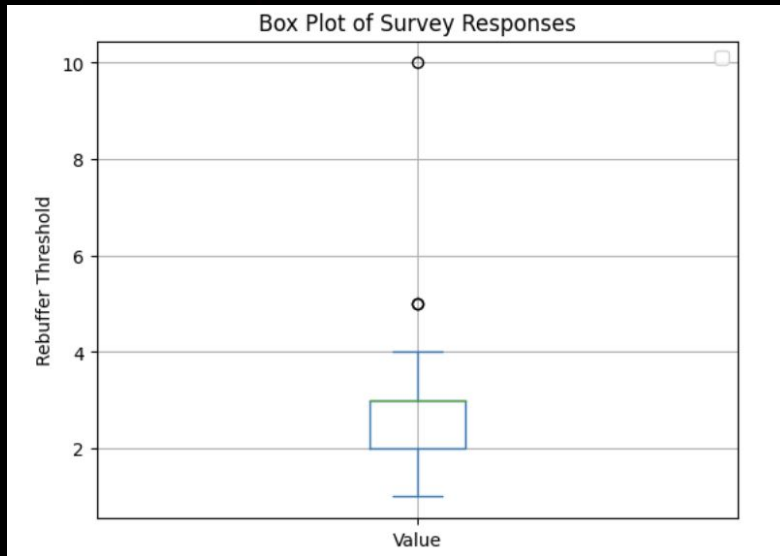
... two rebuffers?

... three rebuffers?

... four rebuffers? ...

Cumulative Distribution Function

Describes the probability that a random variable takes on a value less than or equal to a specific value. It provides a complete description of the distribution.



Level 3: Implementing Surveys



Pro: Relatively lightweight

Con: Potentially skewed by the population we polled

Level 3: Implementing Surveys



Leveraging in app surveys

**How was your experience playing
Underwatermelon: Fruit Merge?**

1 - Poor 2 - Fair 3 - OK 4 - Good 5 - Excellent

Skip

What could we improve?

Video quality Delayed input

Audio quality Gameplay quality

Other

Optional additional information

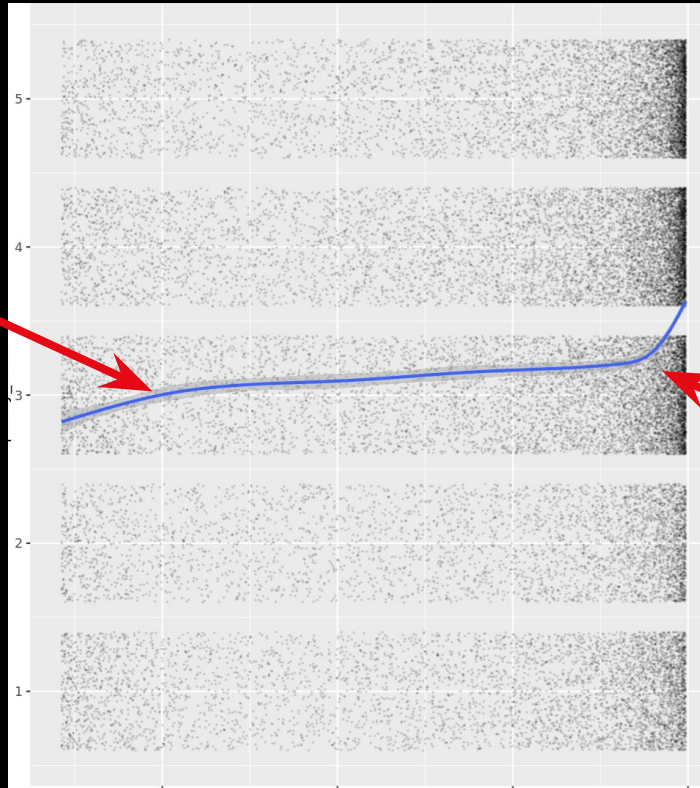
Continue

Level 3: Implementing Surveys



Define points
of interest

Survey Rating



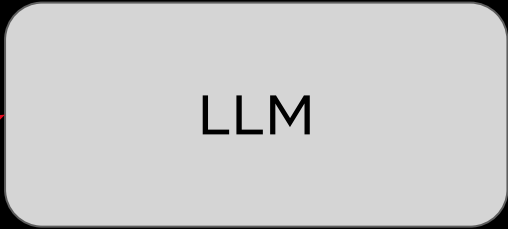
Level 3: Implementing Surveys



What could we improve?

<input type="checkbox"/> Video quality	<input type="checkbox"/> Delayed input
<input type="checkbox"/> Audio quality	<input type="checkbox"/> Gameplay quality
<input type="checkbox"/> Other	
<input type="text" value="Optional additional information"/>	

Continue



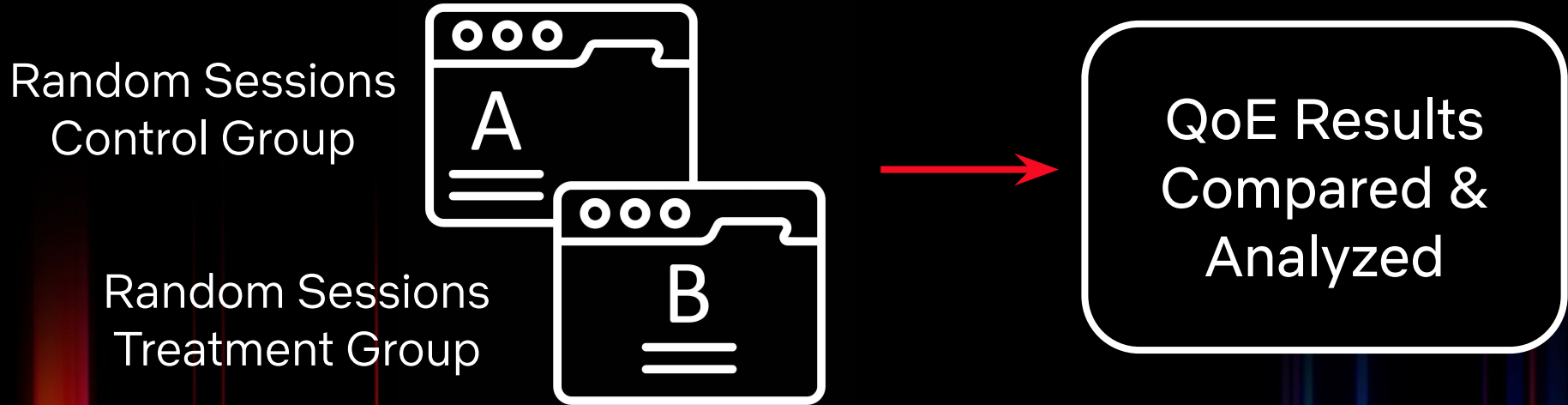
Level 3: Implementing Surveys



Pro: Poll the population of users

Con: People's rating of a service may not always correlate with their behavior

Level 3: Leveraging AB Testing

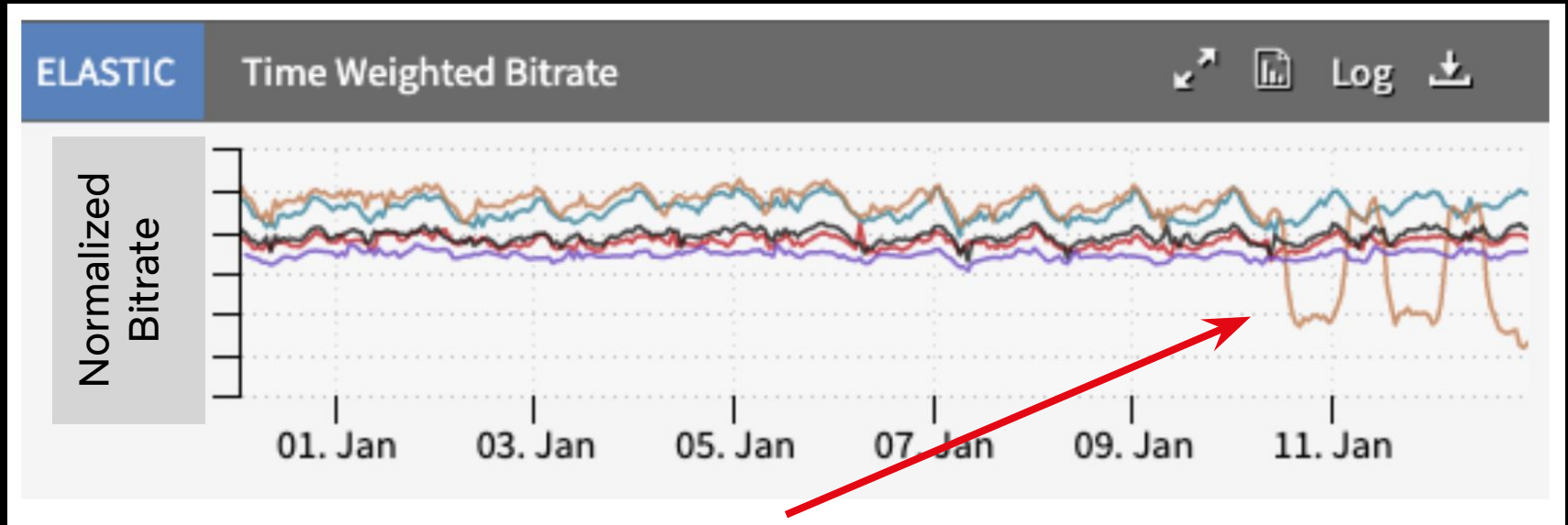


Pro: Diverse Conditions with Real Users from the Production System

Episode 4

Actioning on Alerts

QoE Regression Detection






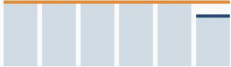





congested data center site

Costa Rica

Leaderboard

Last Updated: August 2024

Rank	Speed (Mbps)	ISP ⓘ	Type	Last 6 Months
1	 3.0	Claro	Fiber	
2	 2.8	Cabletica	Cable	
		Coopelesca	Fiber	
		Telecable	Fiber	
		Tigo	Cable	
3	 2.6	Kolbi	Cable DSL	

Episode 5

Take Away

SREs are incredibly impactful to the end product, we are delivering the content to millions of people.

Understanding high level concepts in data science is a super power.



Thank You.



Thiara Ortiz
Staff CDN Reliability Engineer