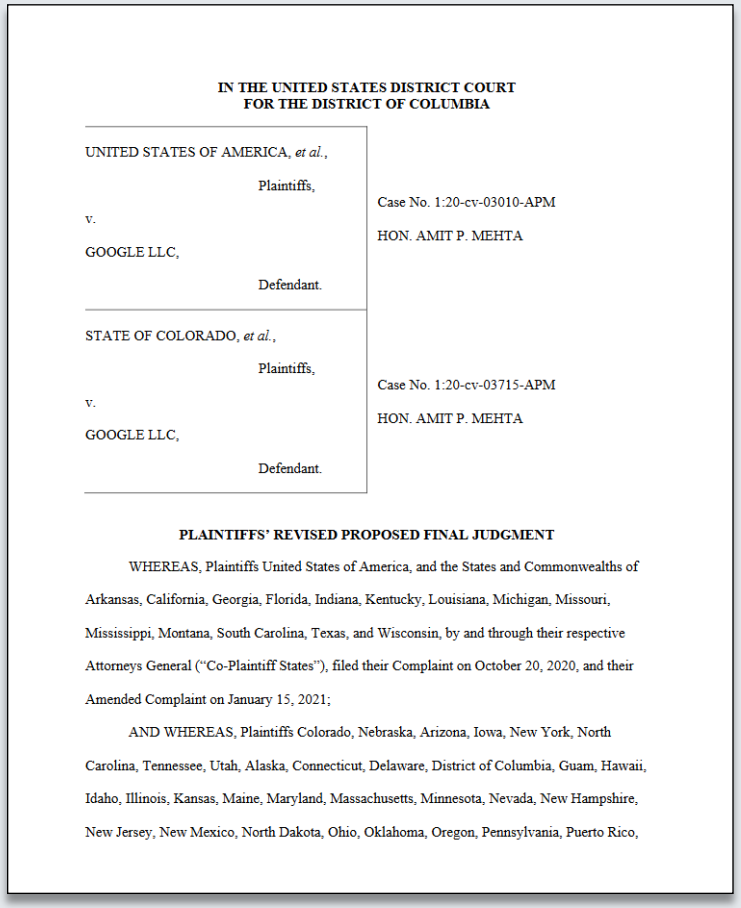
Prof. James Mickens

Distributed Systems and Security Expert

PUBLIC REDACTED VERSION

Assignment



IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

UNITED STATES OF AMERICA, *et al.*,
Plaintiffs,
v.
GOOGLE LLC,
Defendant.
Case No. 1:20-cv-03010-APM
HON. AMIT P. MEHTA

STATE OF COLORADO, *et al.*,
Plaintiffs,
v.
GOOGLE LLC,
Defendant.
Case No. 1:20-cv-03715-APM
HON. AMIT P. MEHTA

PLAINTIFFS' REVISED PROPOSED FINAL JUDGMENT

WHEREAS, Plaintiffs United States of America, and the States and Commonwealths of Arkansas, California, Georgia, Florida, Indiana, Kentucky, Louisiana, Michigan, Missouri, Mississippi, Montana, South Carolina, Texas, and Wisconsin, by and through their respective Attorneys General ("Co-Plaintiff States"), filed their Complaint on October 20, 2020, and their Amended Complaint on January 15, 2021;

AND WHEREAS, Plaintiffs Colorado, Nebraska, Arizona, Iowa, New York, North Carolina, Tennessee, Utah, Alaska, Connecticut, Delaware, District of Columbia, Guam, Hawaii, Idaho, Illinois, Kansas, Maine, Maryland, Massachusetts, Minnesota, Nevada, New Hampshire, New Jersey, New Mexico, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Puerto Rico,

Assignment 1: Assess divesting Chrome from a technical perspective

Assignment 2: Assess data sharing and syndication remedies from a technical perspective

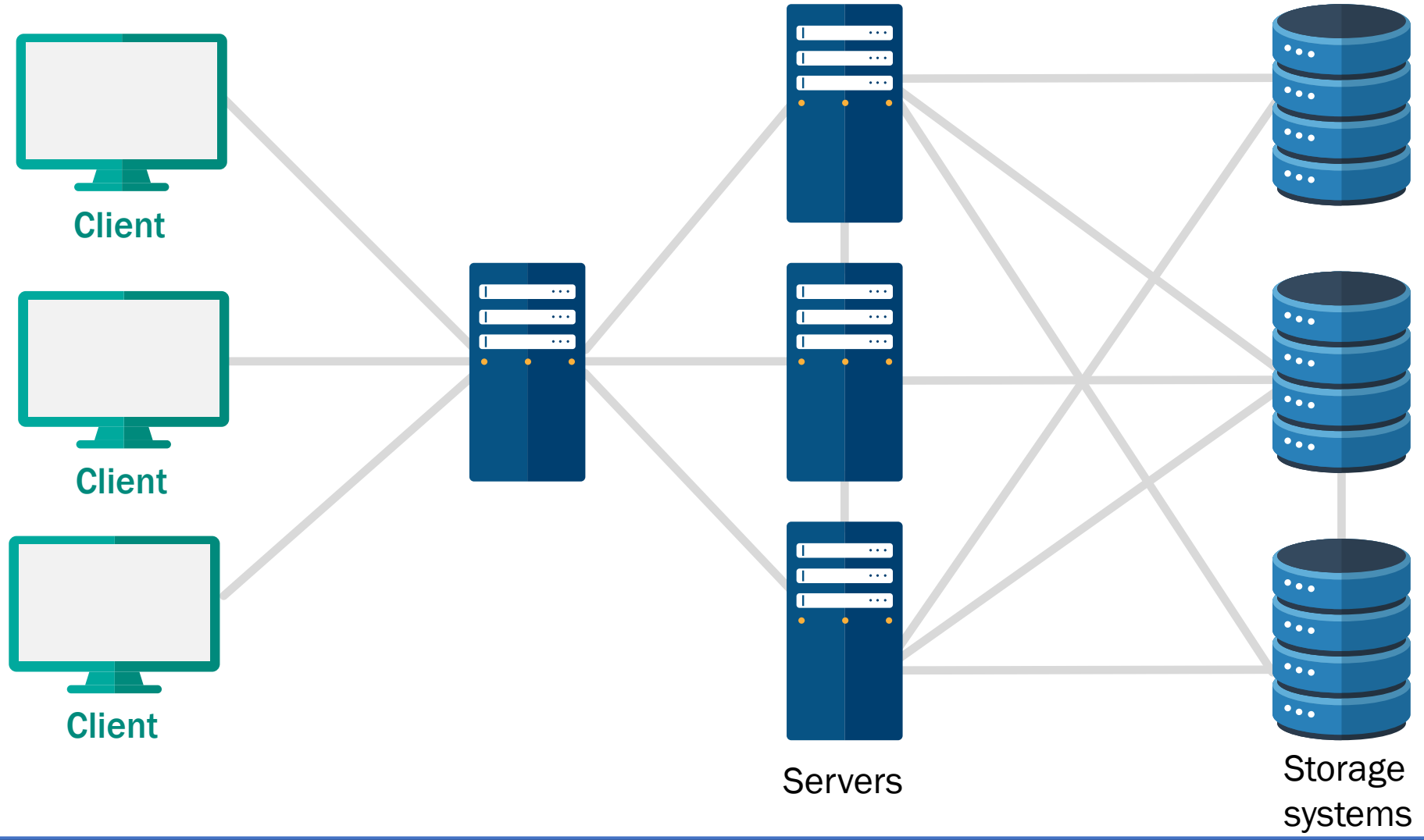
Why Distributed Systems?

Understanding distributed systems is a necessary step in understanding the divestiture, data sharing, and syndication remedies.

Distributed Systems: An Overview

Single Computer

Distributed Systems: An Overview



Key Conclusions

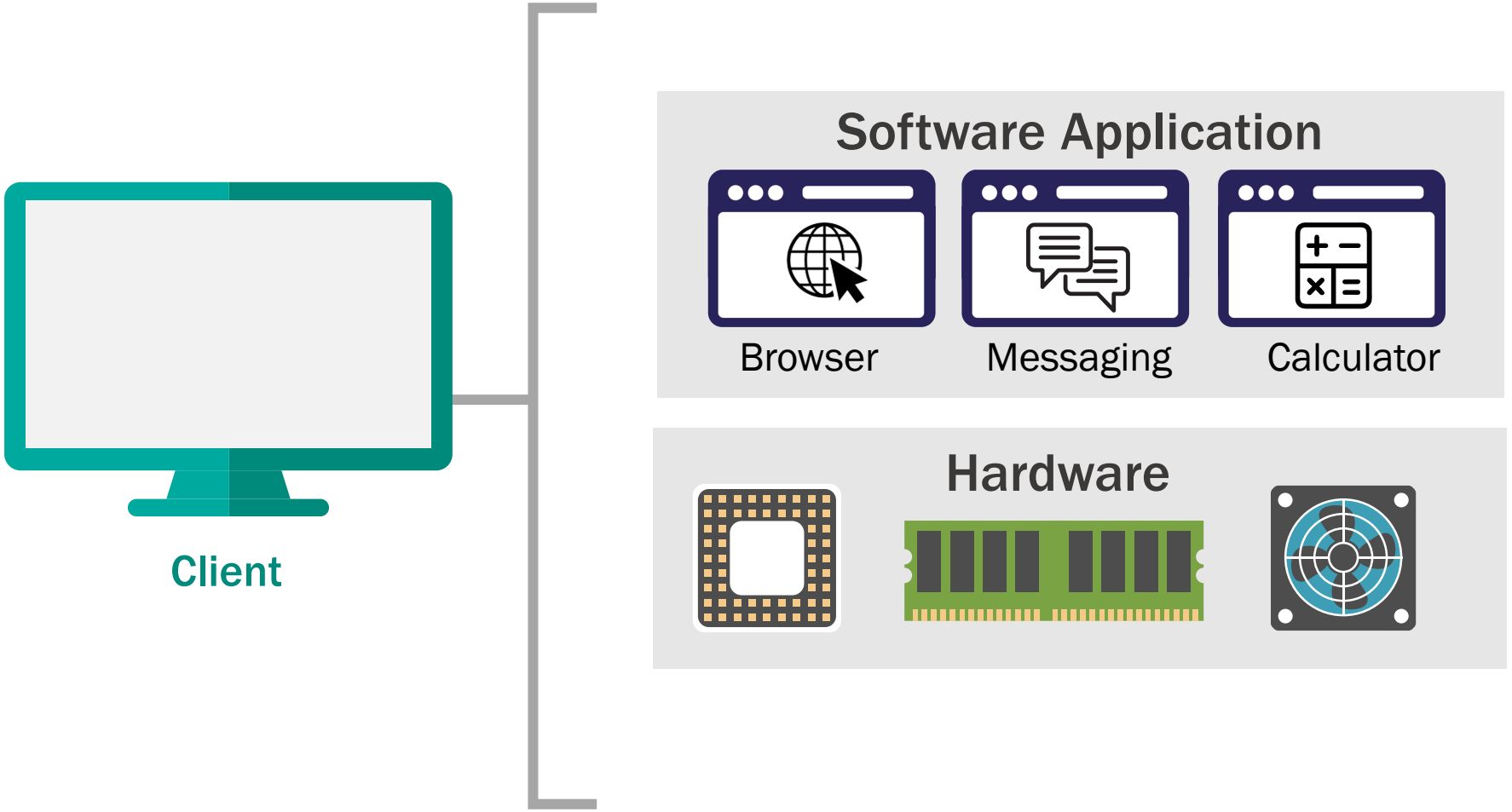
Conclusion 1: From a technical perspective, Chrome divestiture is feasible.

Conclusion 2: From a technical perspective, data sharing and syndication remedies are feasible.

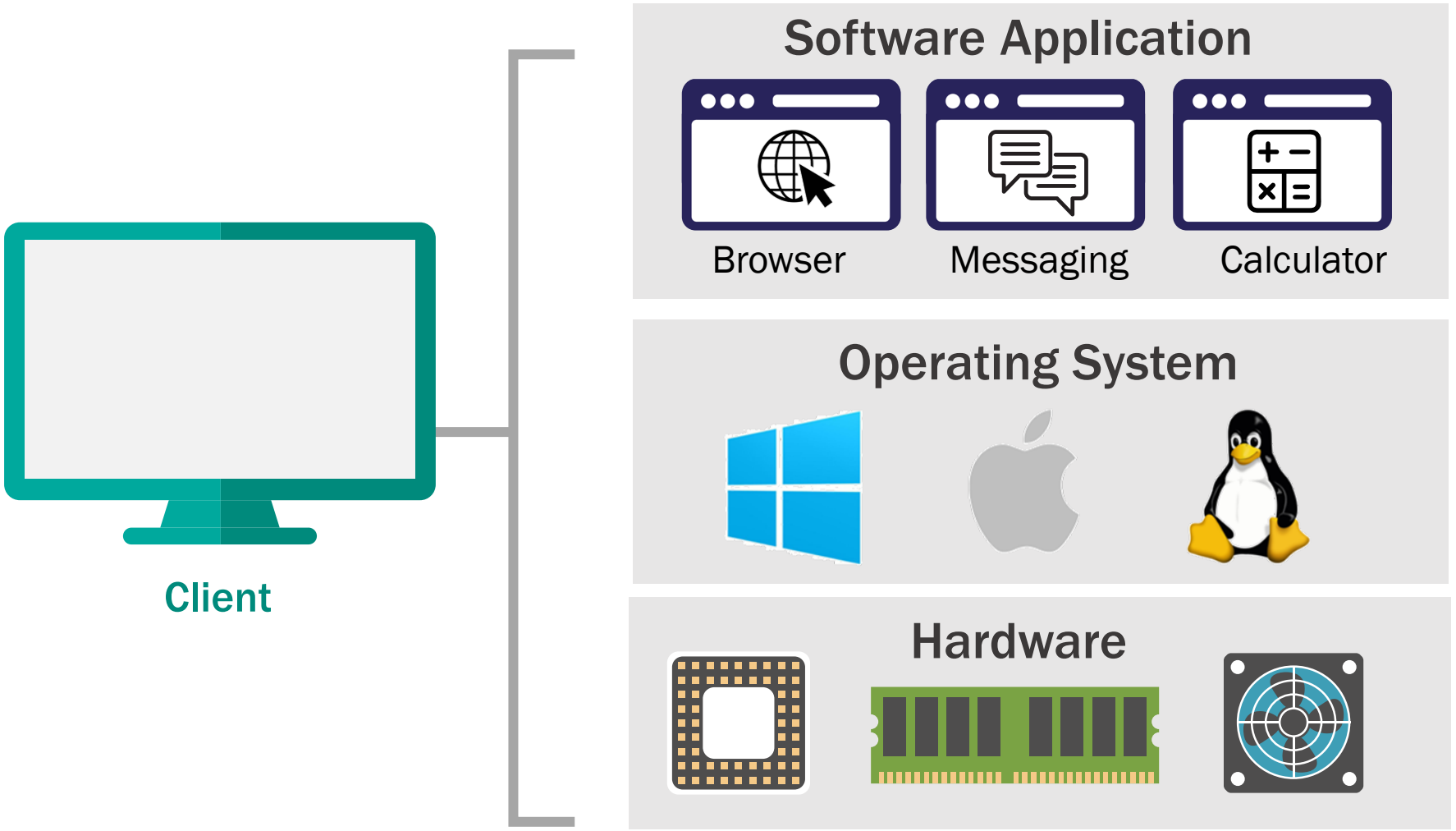
Technical Feasibility

- ☑ Type of work Google and third parties already do.
- ☑ That work is not unduly burdensome.
- ☑ That work can be done in a way that preserves security.

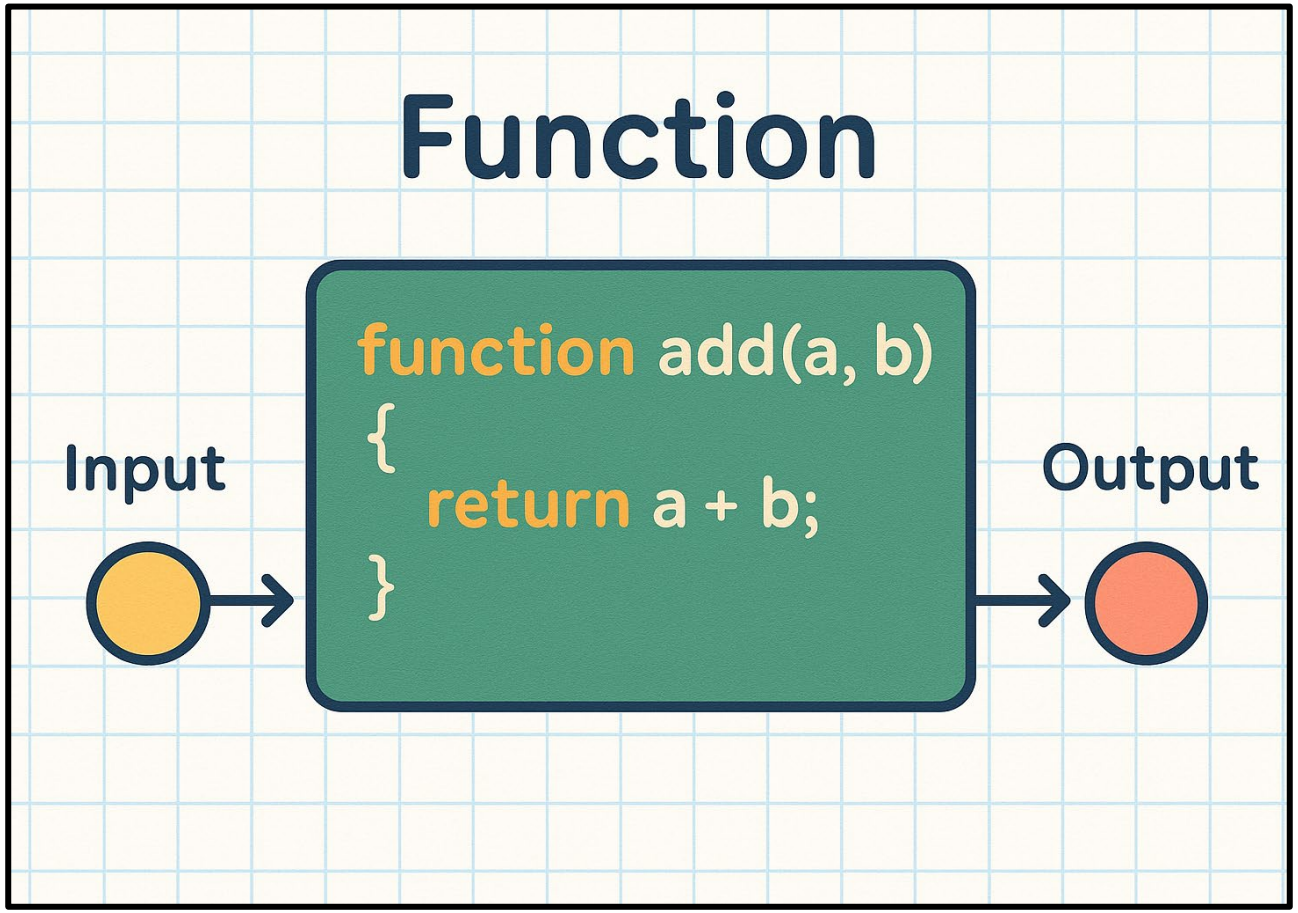
Computers: An Overview



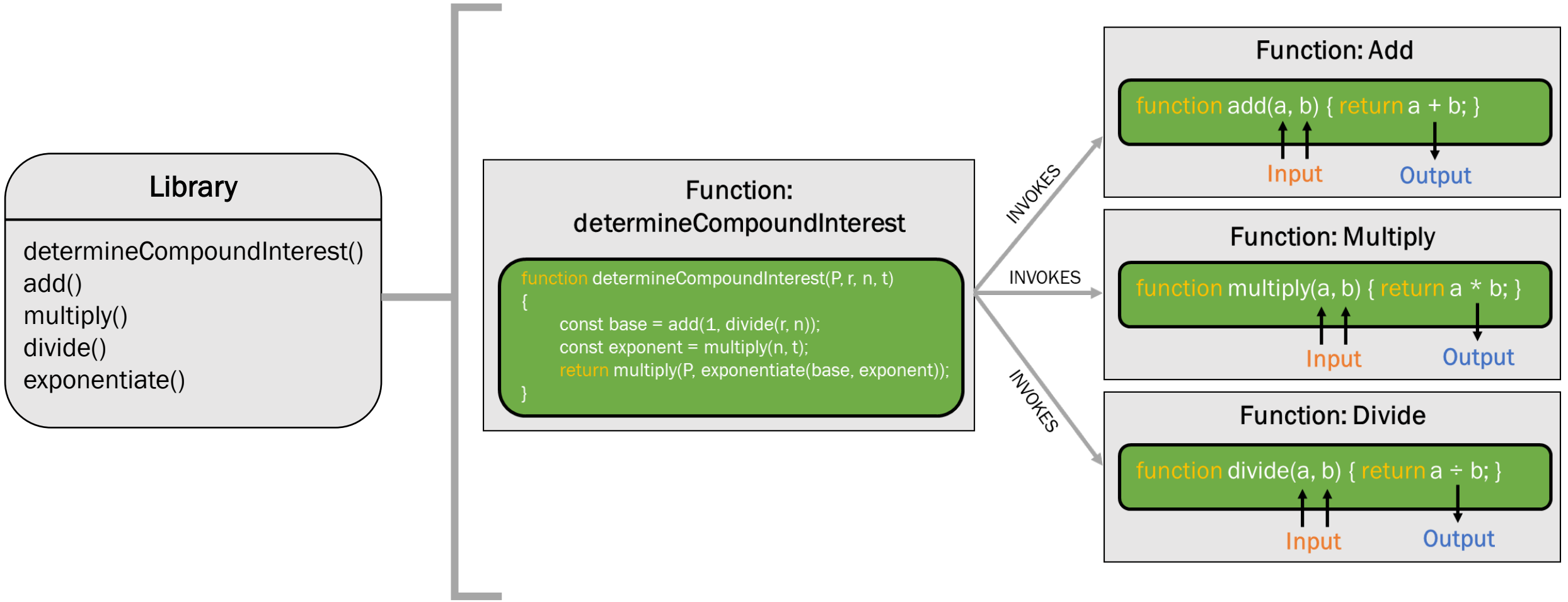
Computers: An Overview



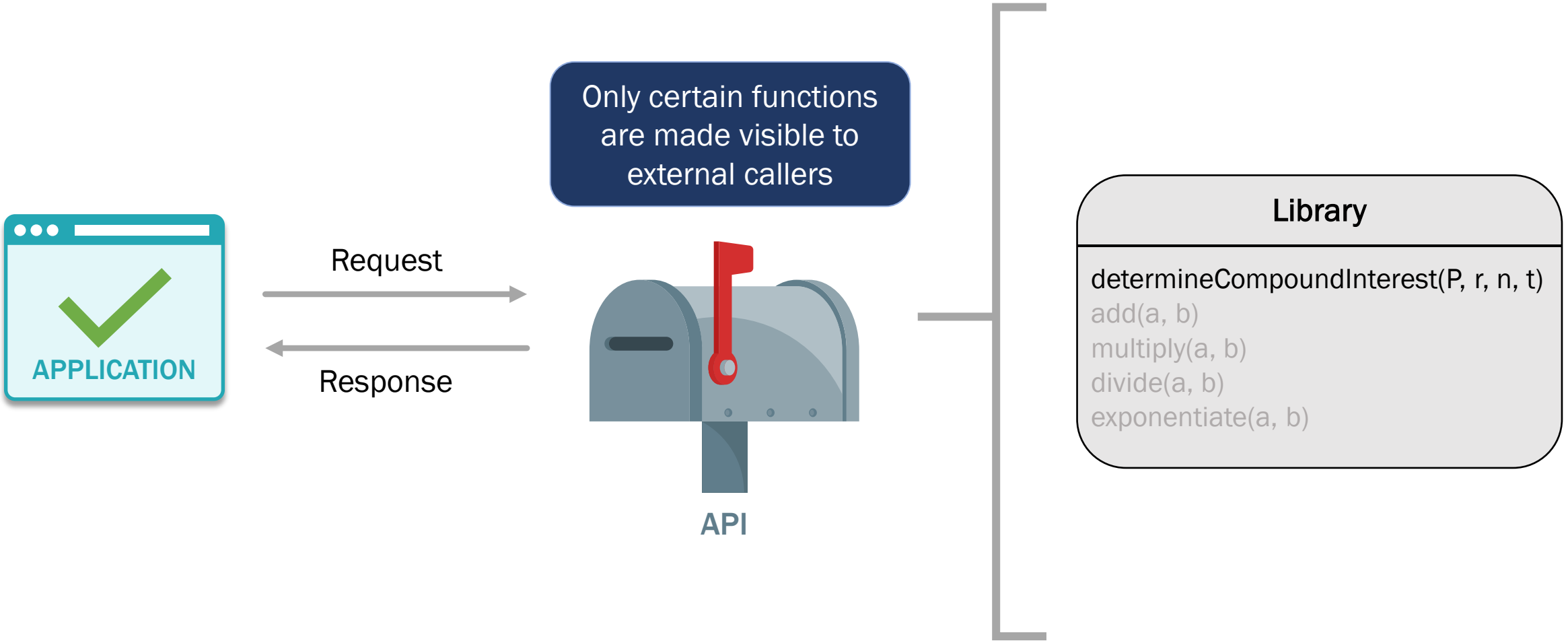
Functions Are Blocks of Code



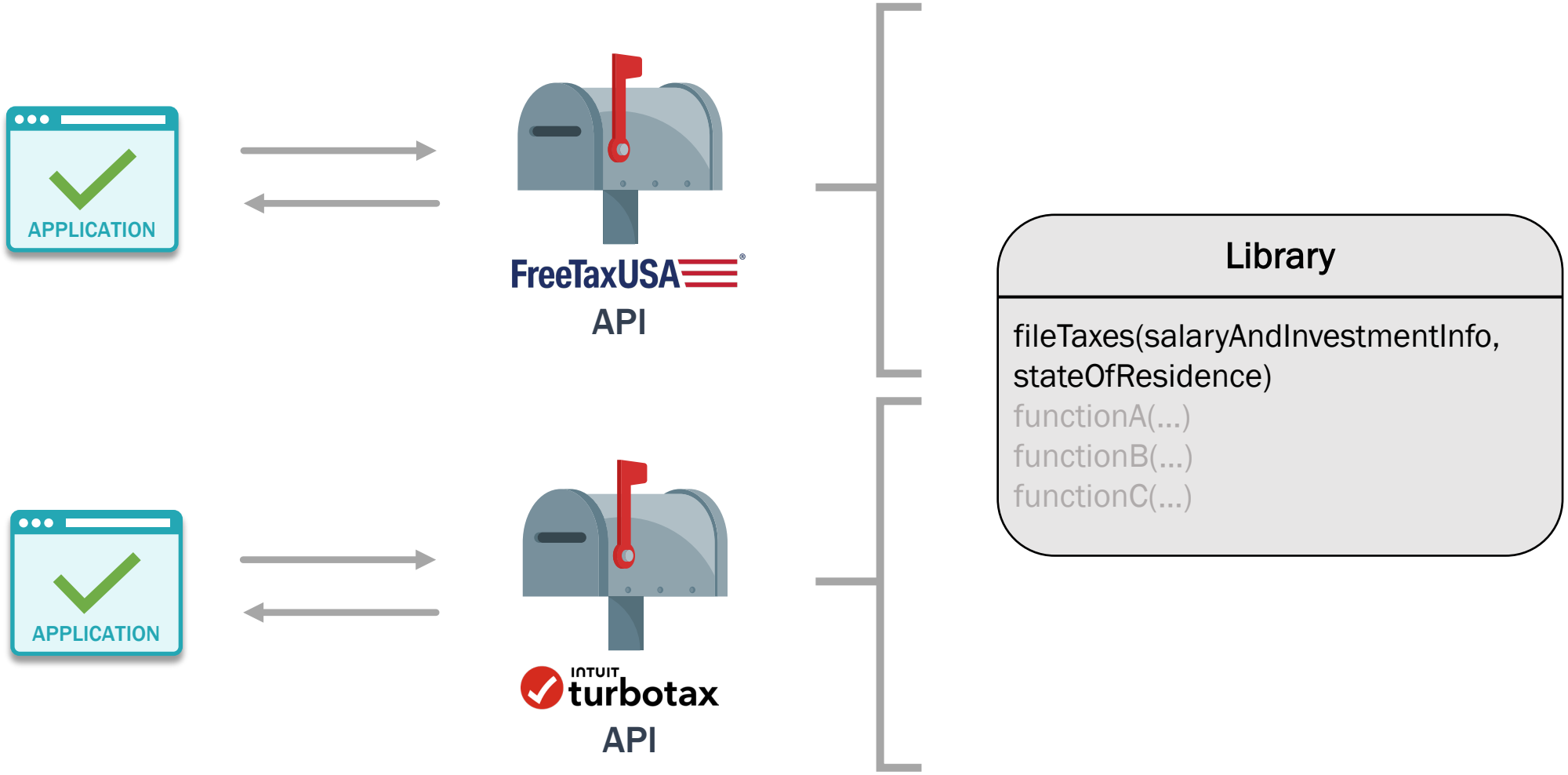
Libraries Are Collections of Functions



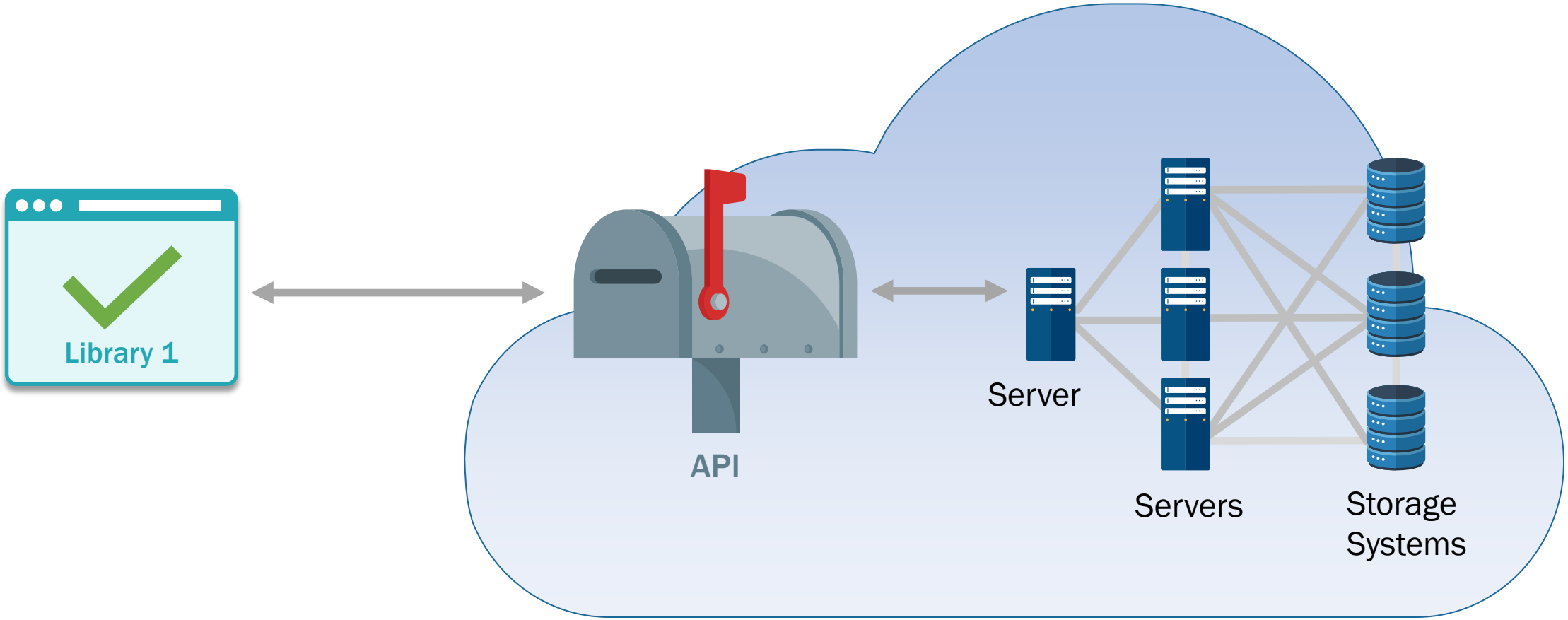
Application Programming Interfaces (APIs) Allow Software Programs to Talk to Each Other



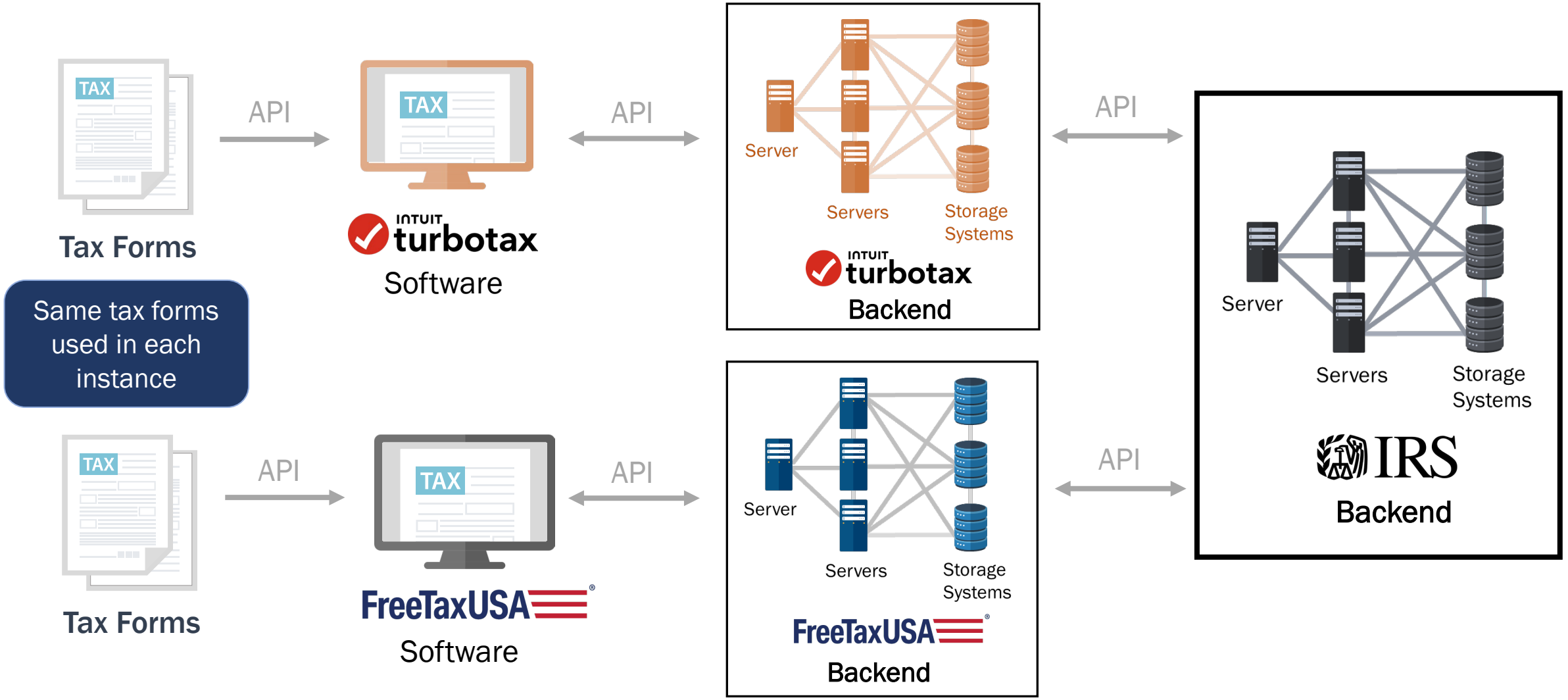
Tax Example: API



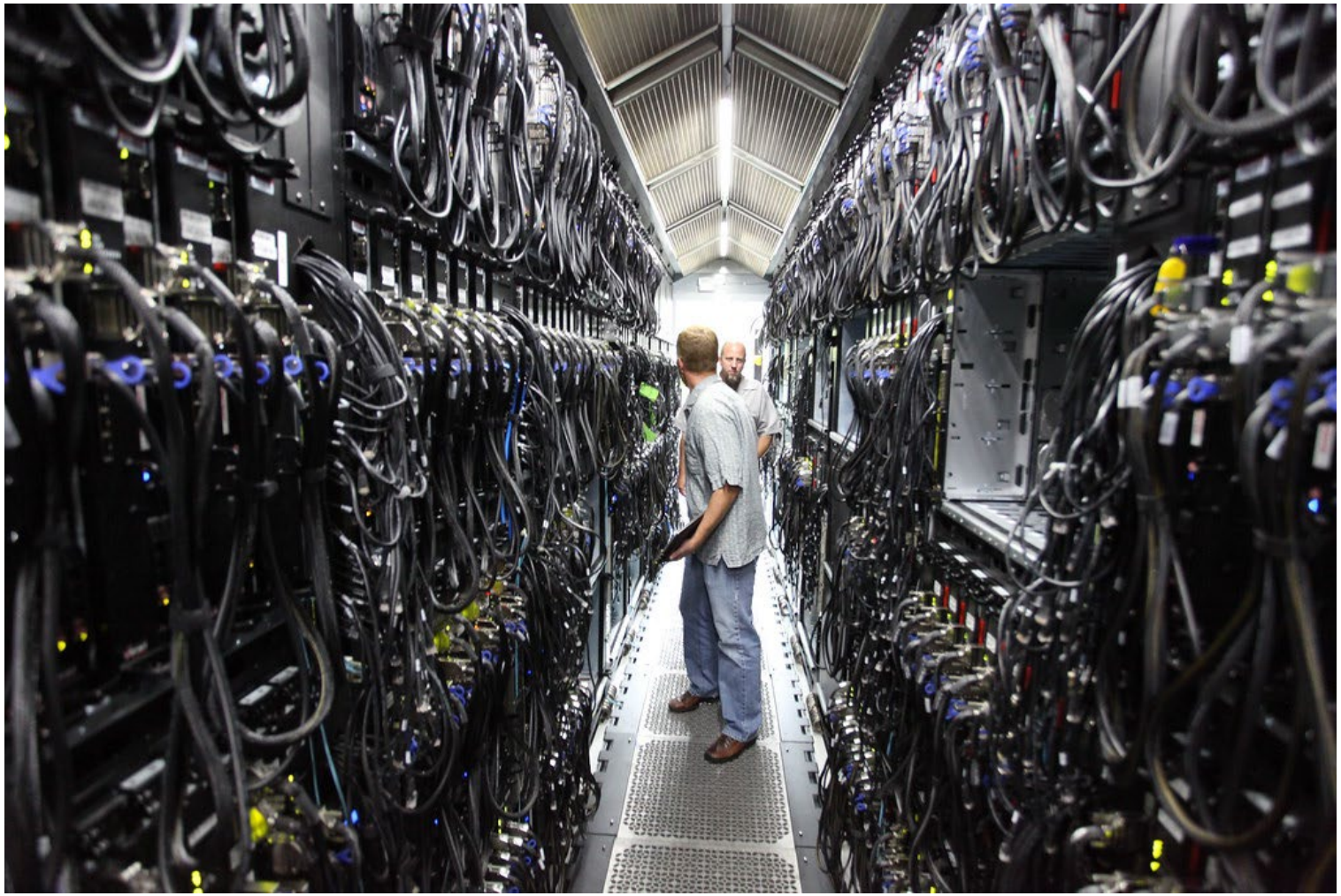
Cloud APIs Connect to Code Running in the Cloud



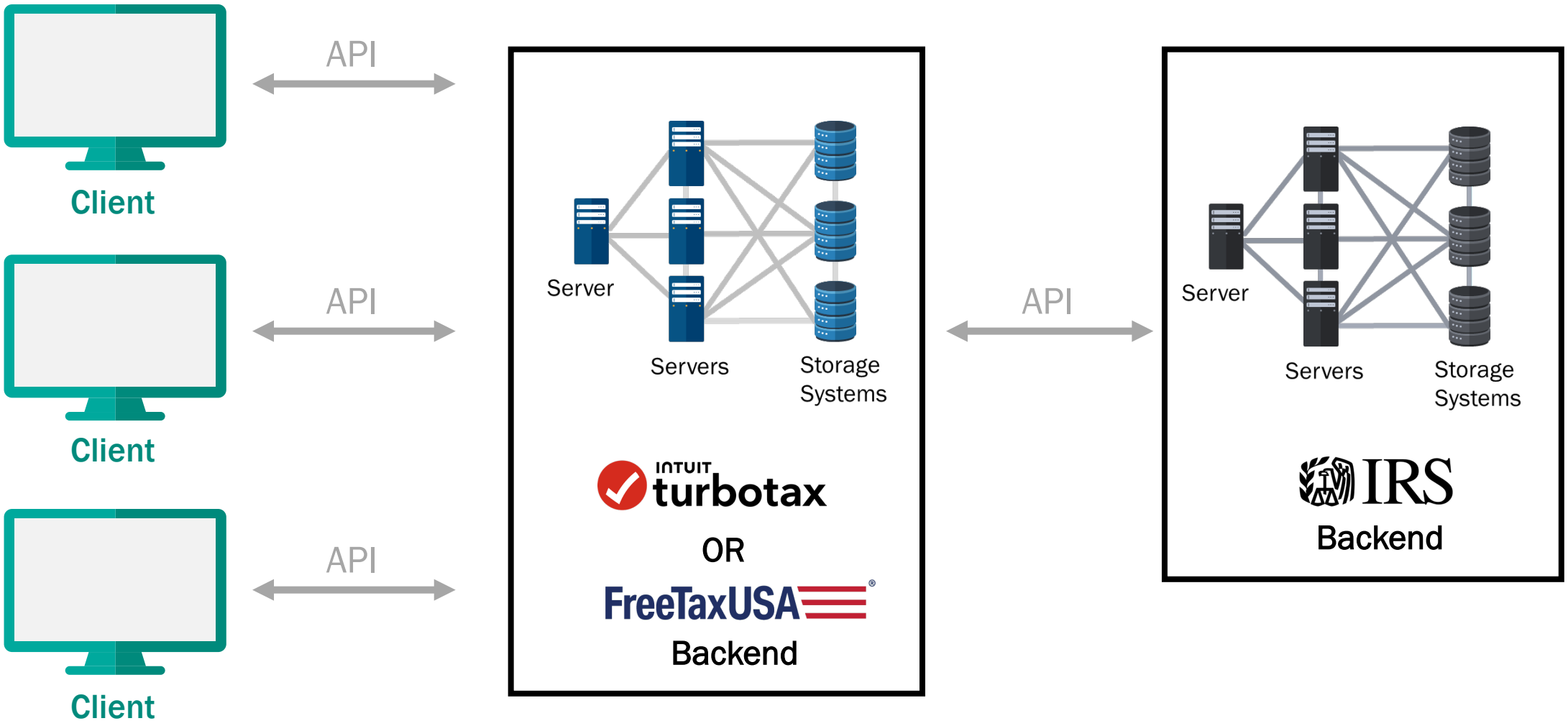
Tax Example: Distributed Systems



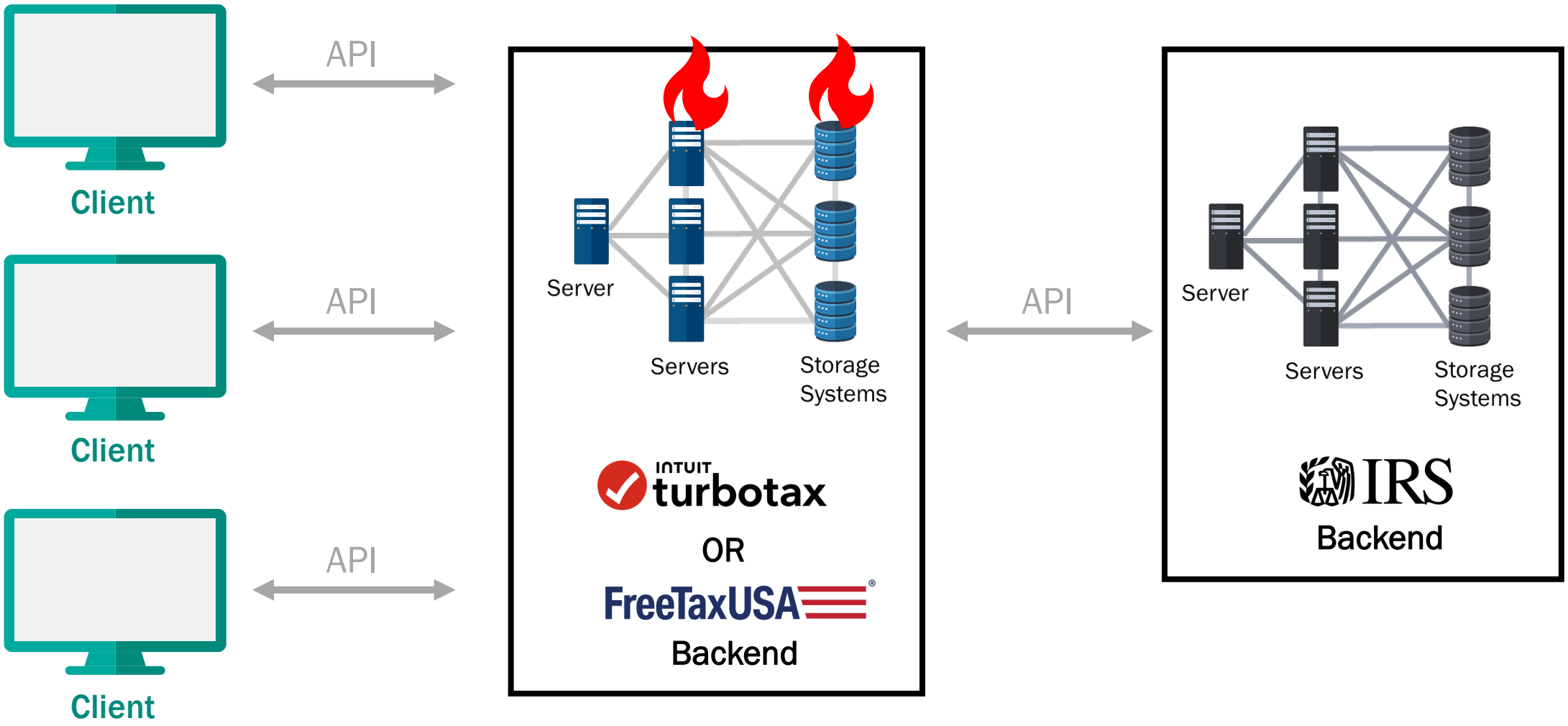
Hardware Infrastructure



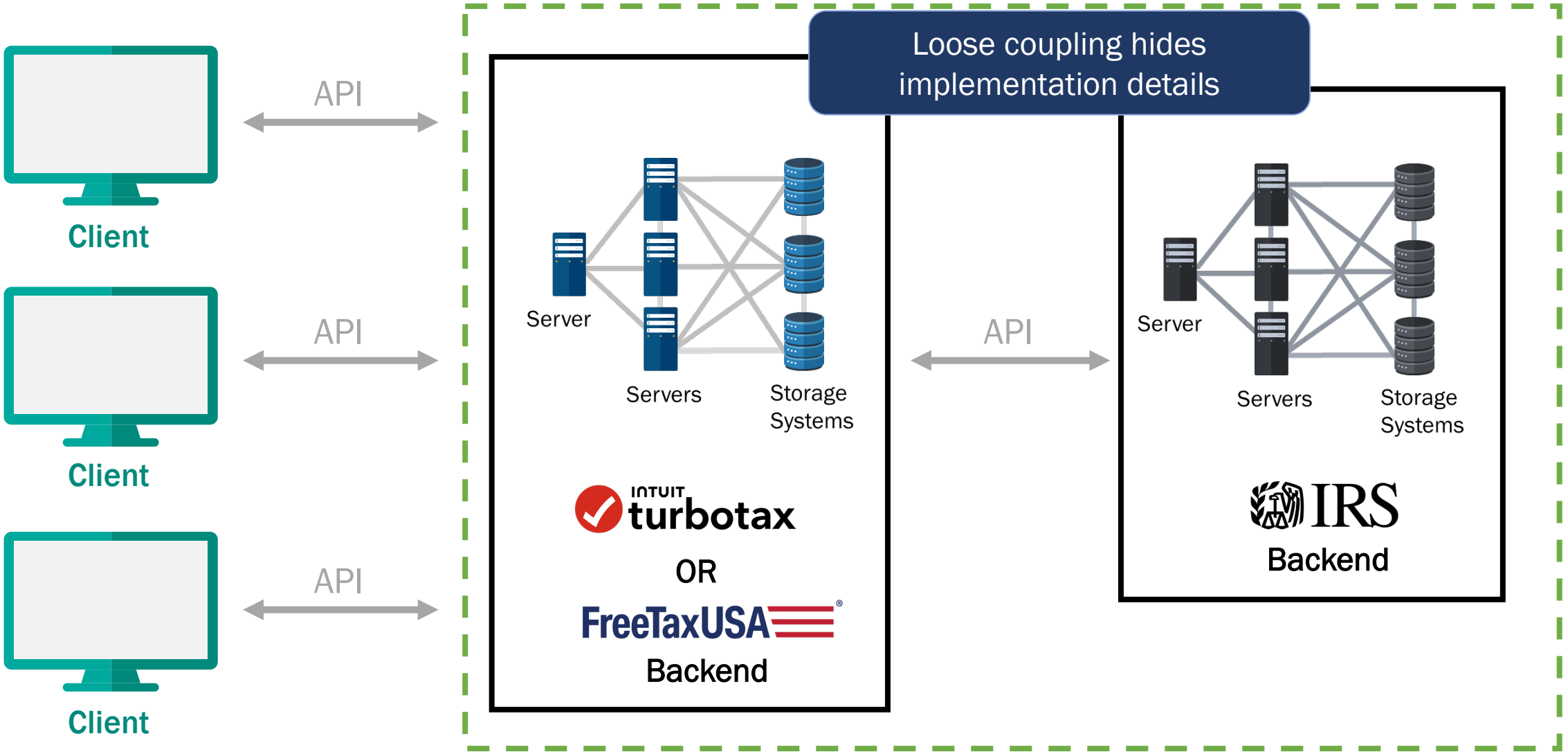
Benefits of Distributed Systems: Scalability and Efficiency



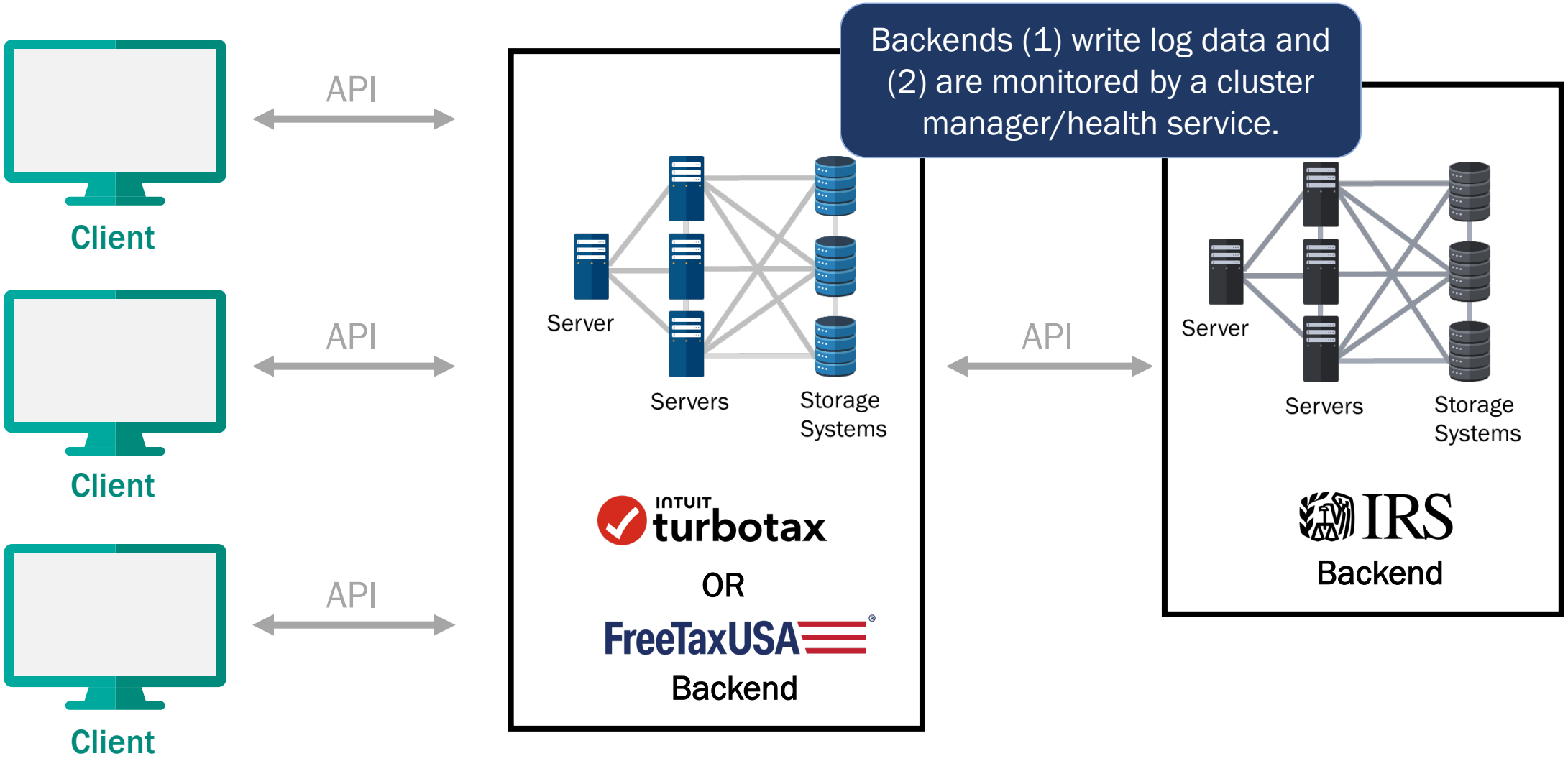
Benefits of Distributed Systems: Reliability



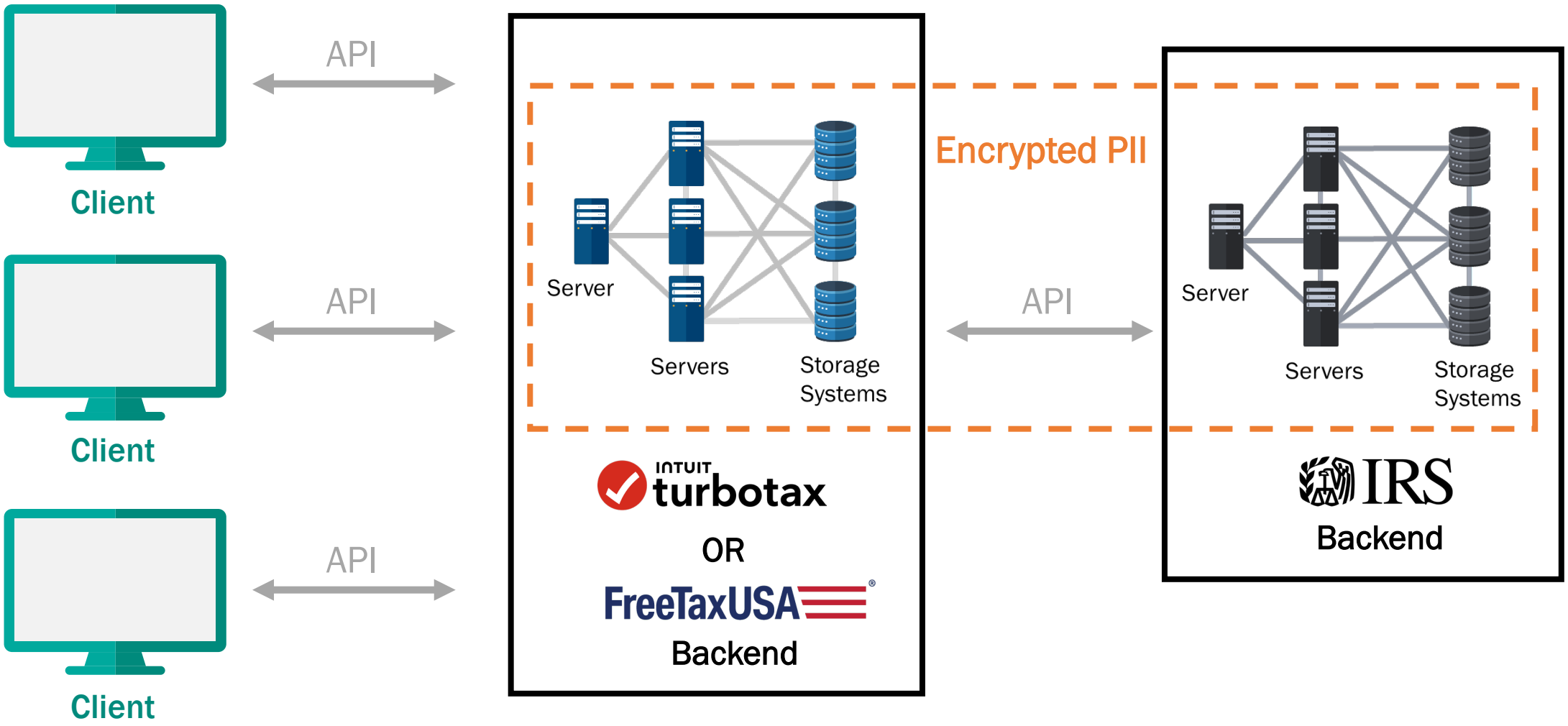
Best Practices with Distributed Systems: Loose Coupling



Best Practices with Distributed Systems: Observability



Best Practices with Distributed Systems: Security

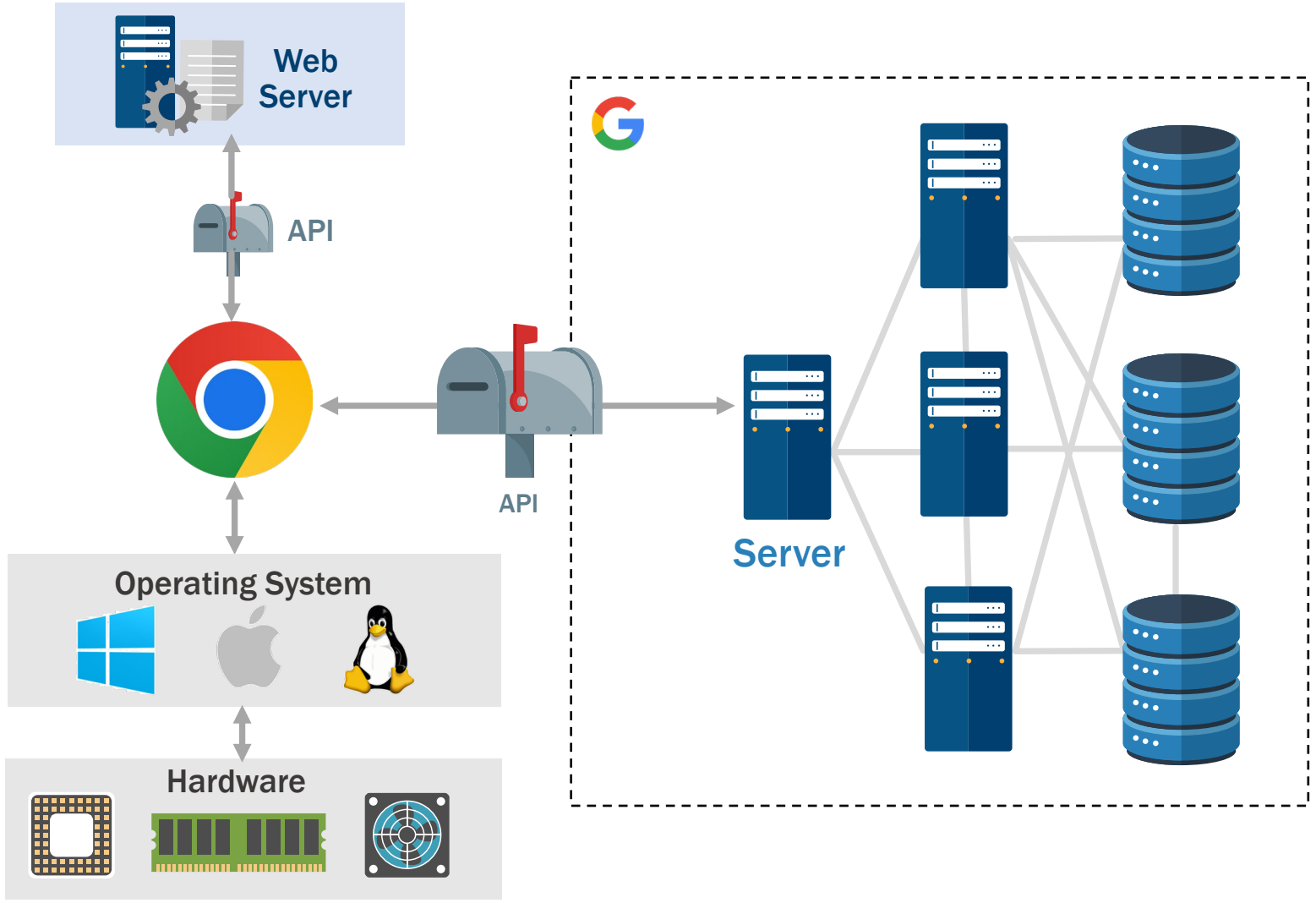


Opinion # 1: Chrome Divestiture is Feasible

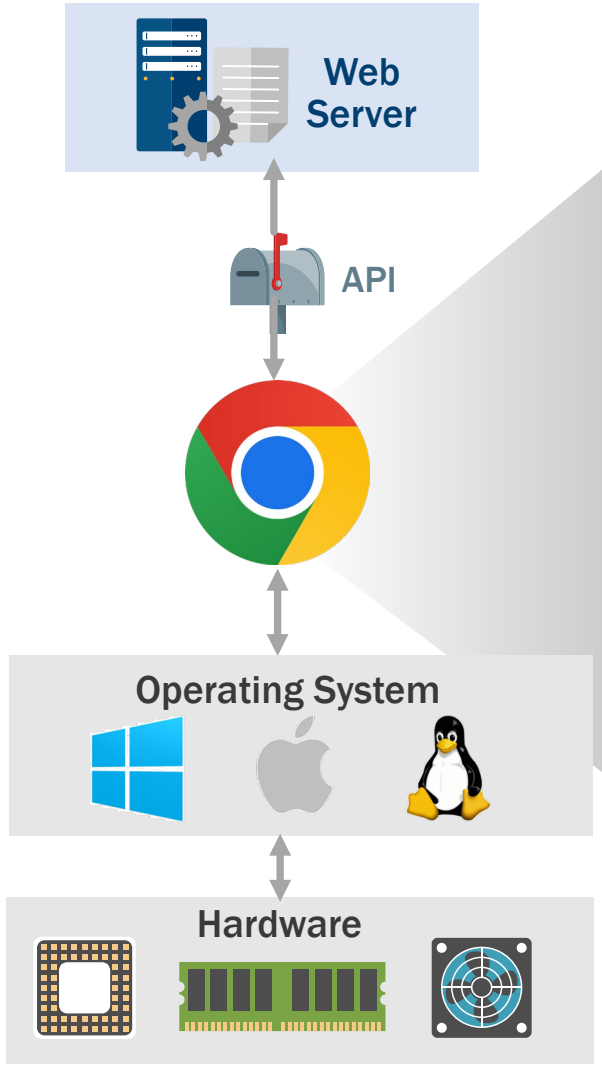
The Chrome browser is not deeply integrated with an underlying operating system.

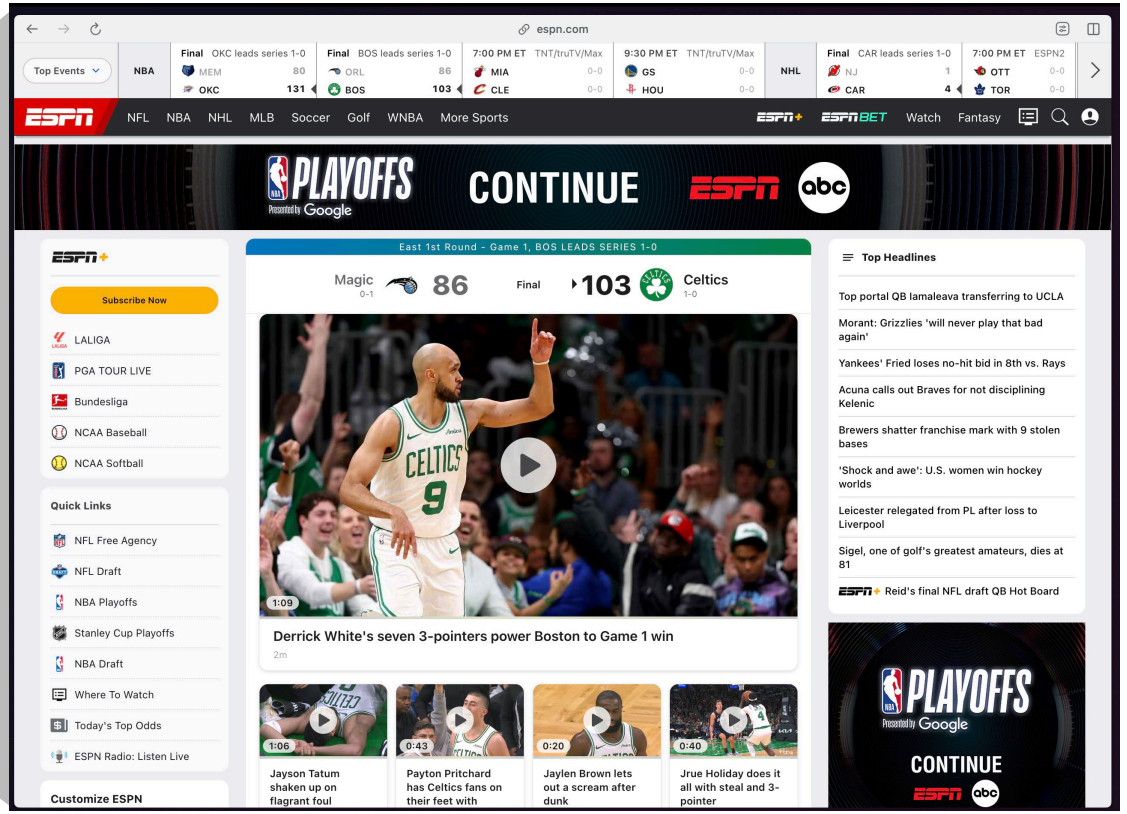
The Chrome browser is not deeply integrated with Google's back-end services.

Chrome is A Piece of Client Software

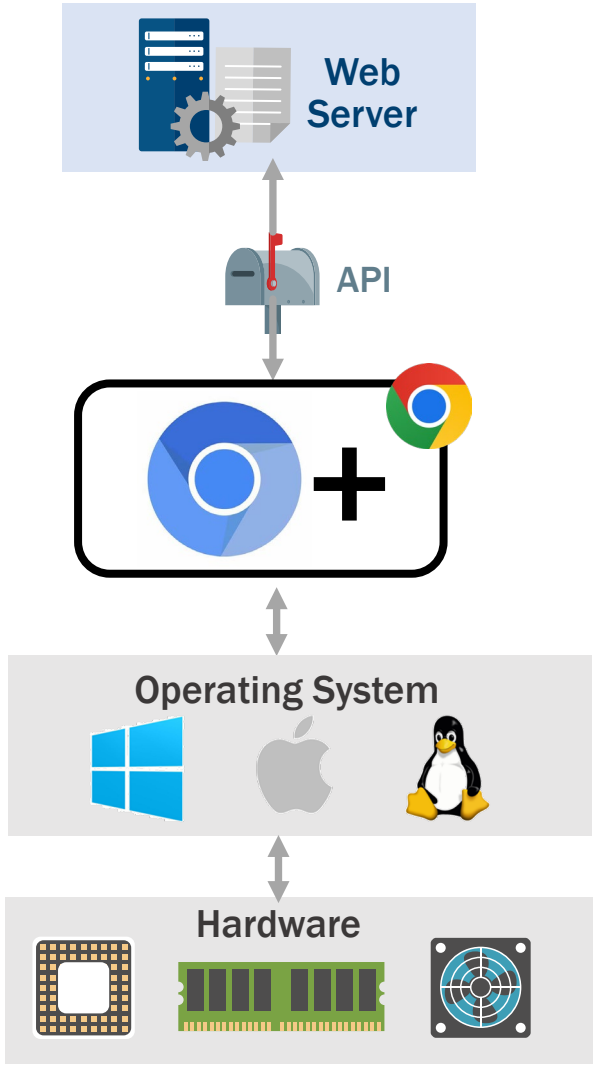


A Browser's Role

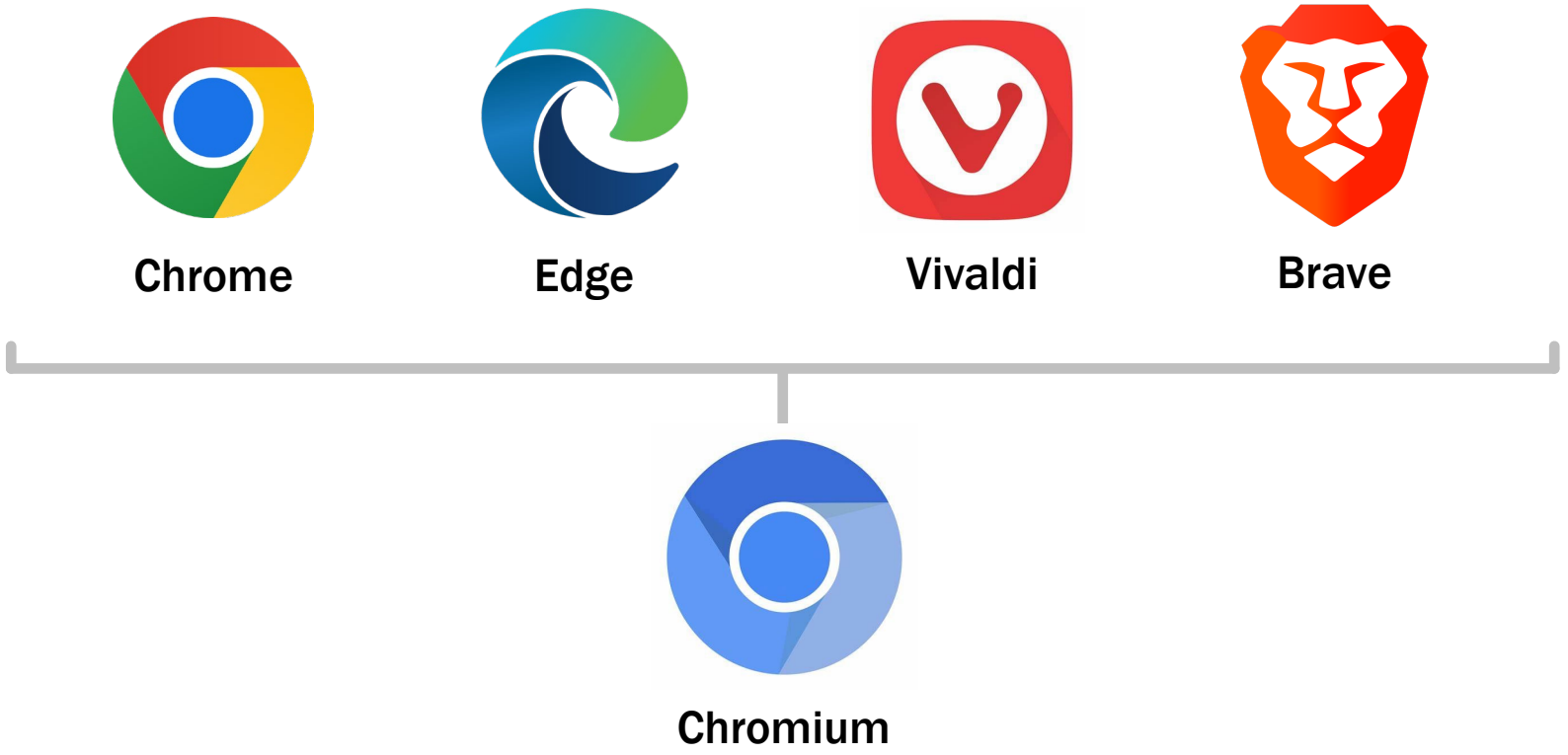




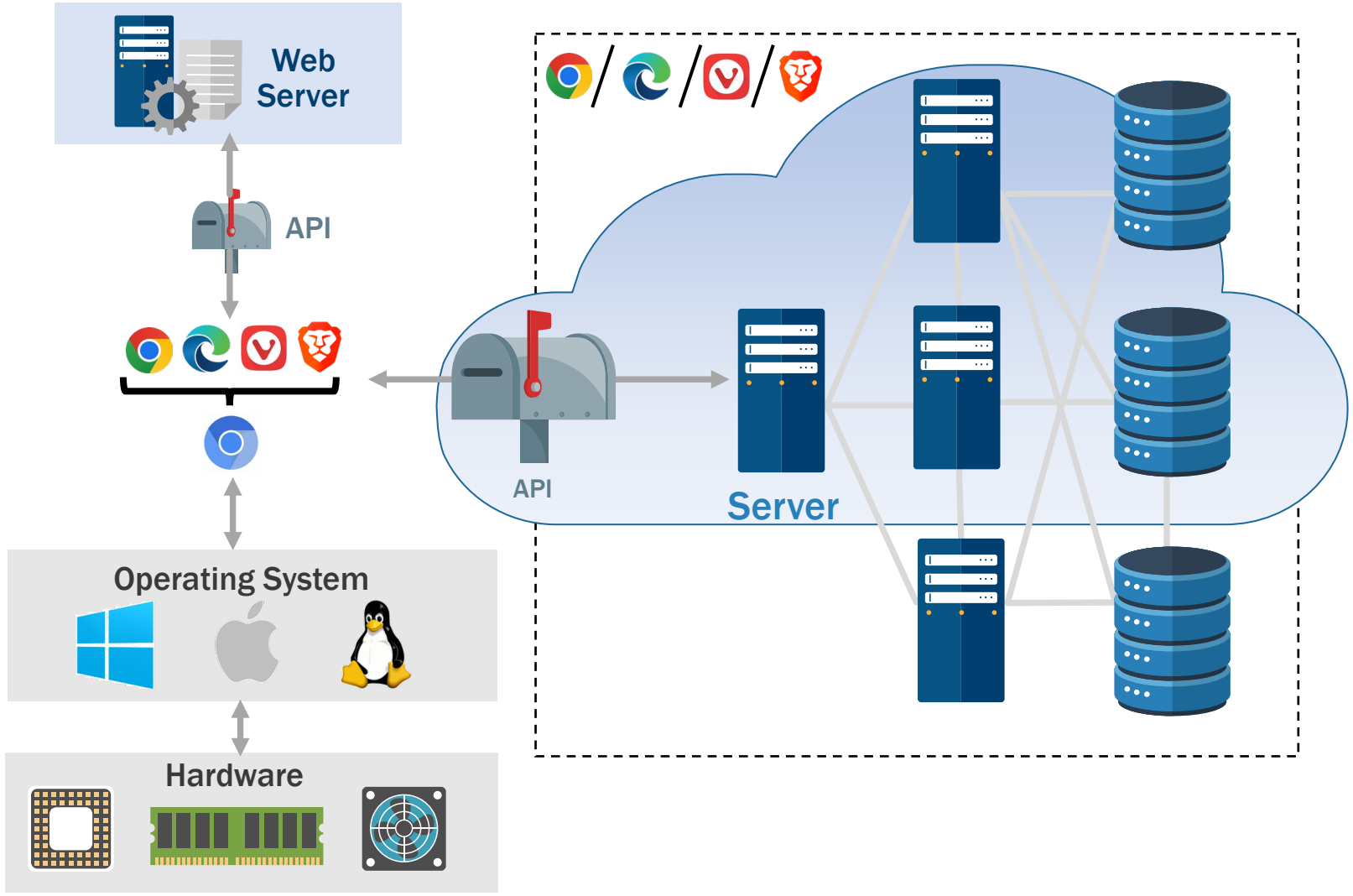
Chrome Is Built on Top of Open-Source Chromium



Other Browsers Are Built on Top of Chromium



Each Chromium-Based Browser Makes Unique Design Decisions



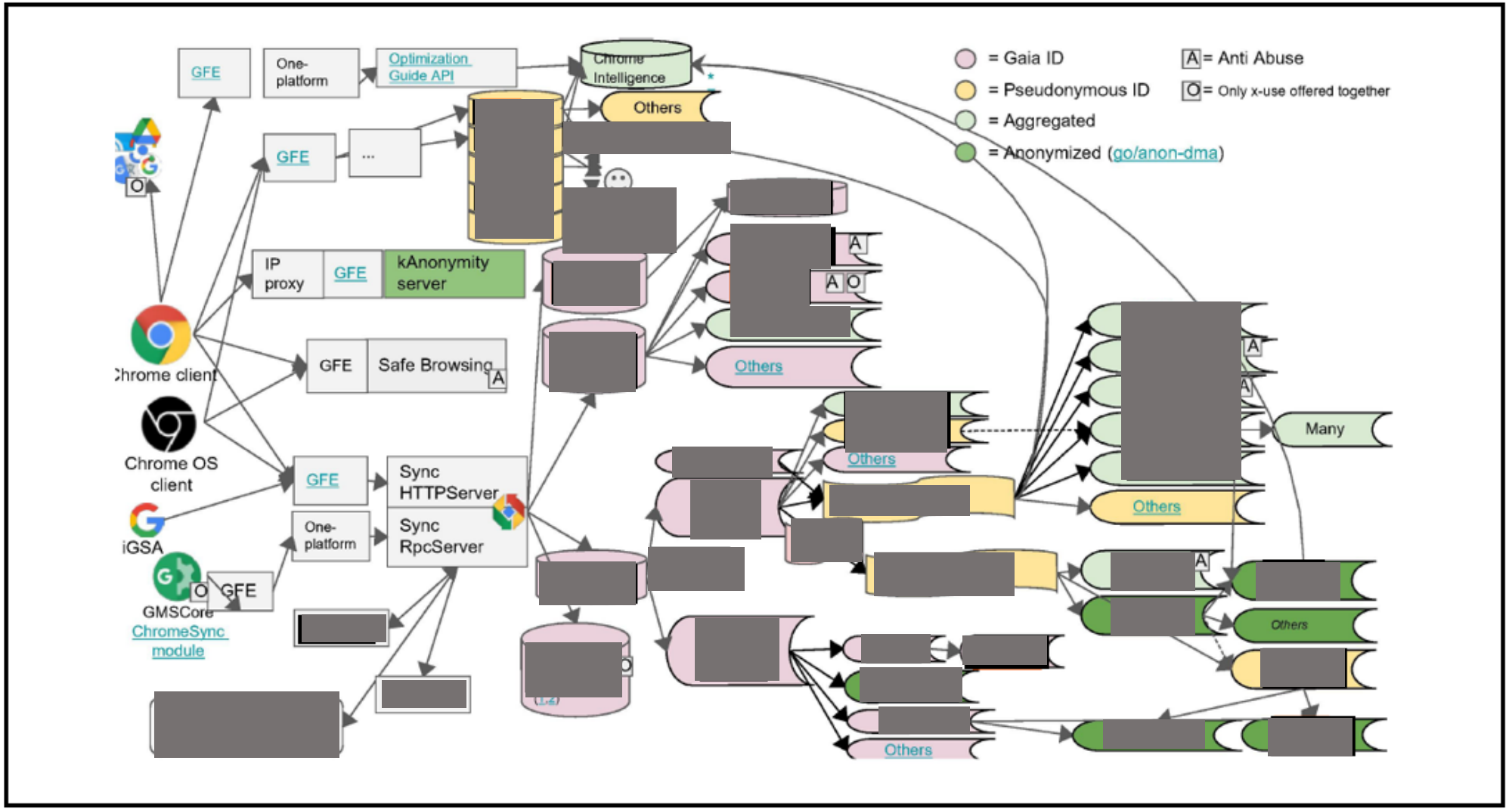
Examples of Google's Backend Services

GAIA API: Authentication service that allows users to log into Chrome.

Chrome Sync: Synchronizes Chrome data across multiple signed-in devices.

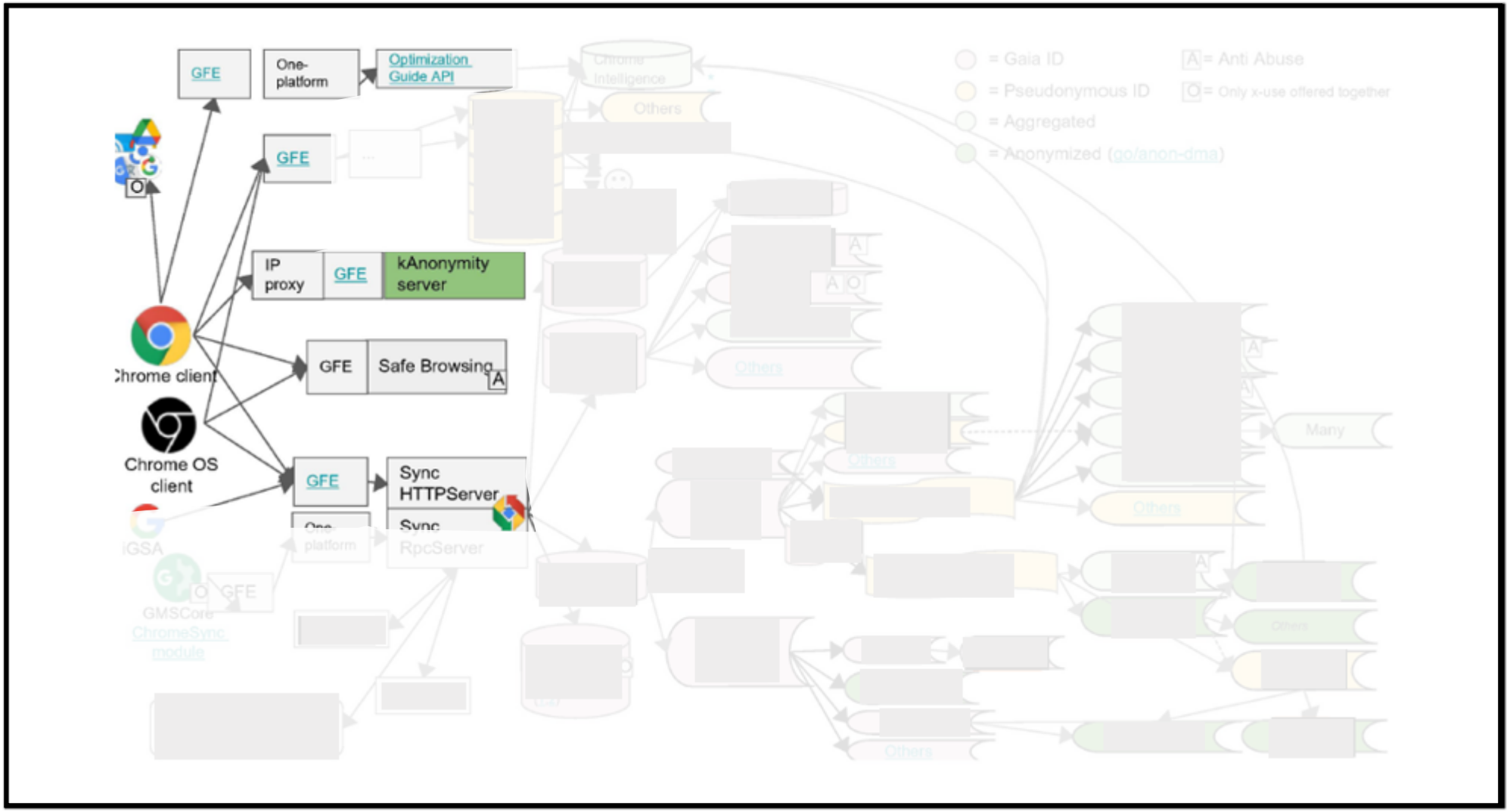
Safe Browsing API: Analyzes and identifies harmful web pages for the Safe Browsing service.

Chrome and Google's Backend



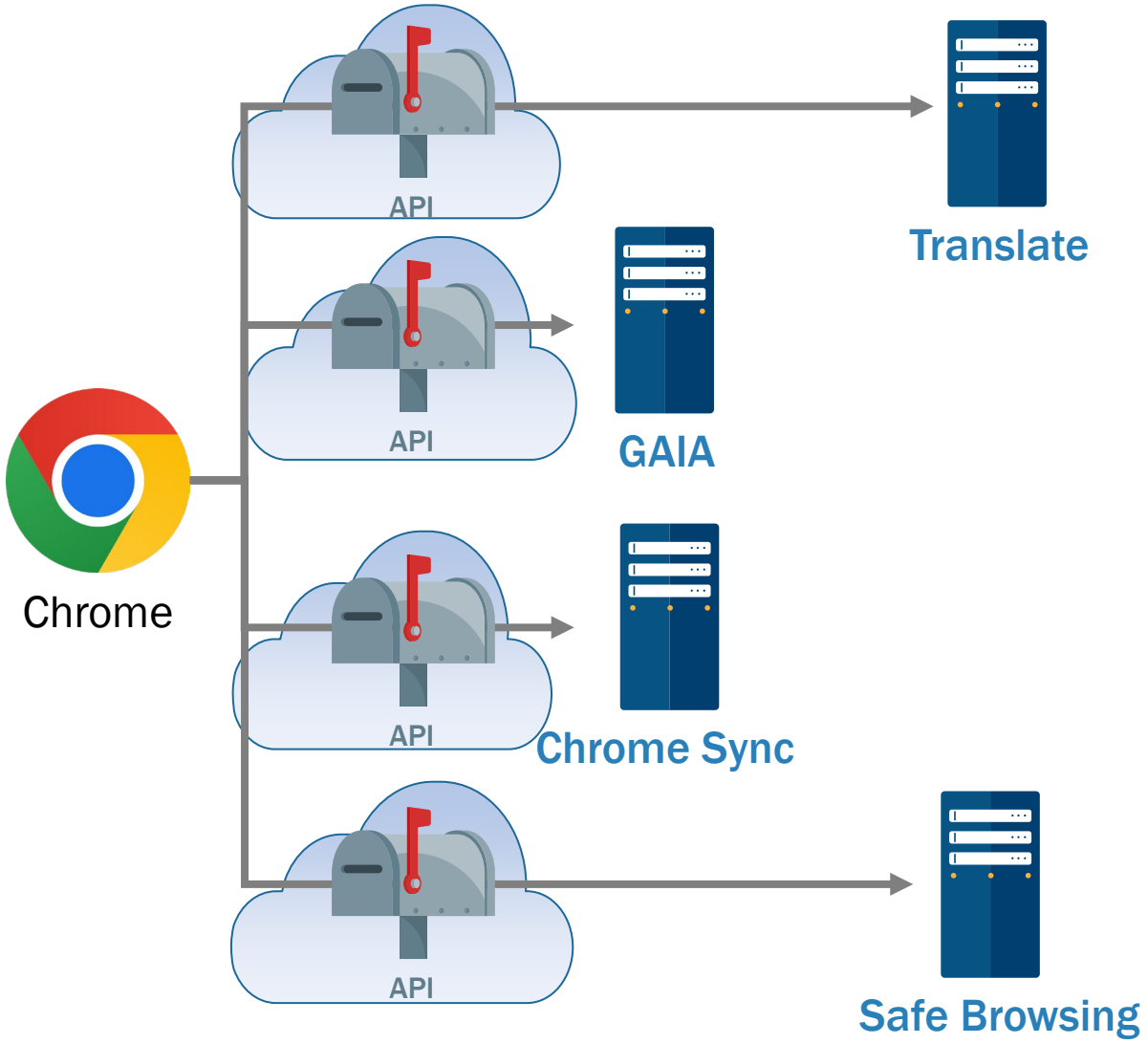
Source: PXR0321

Chrome Doesn't Need to Know Backend Implementation

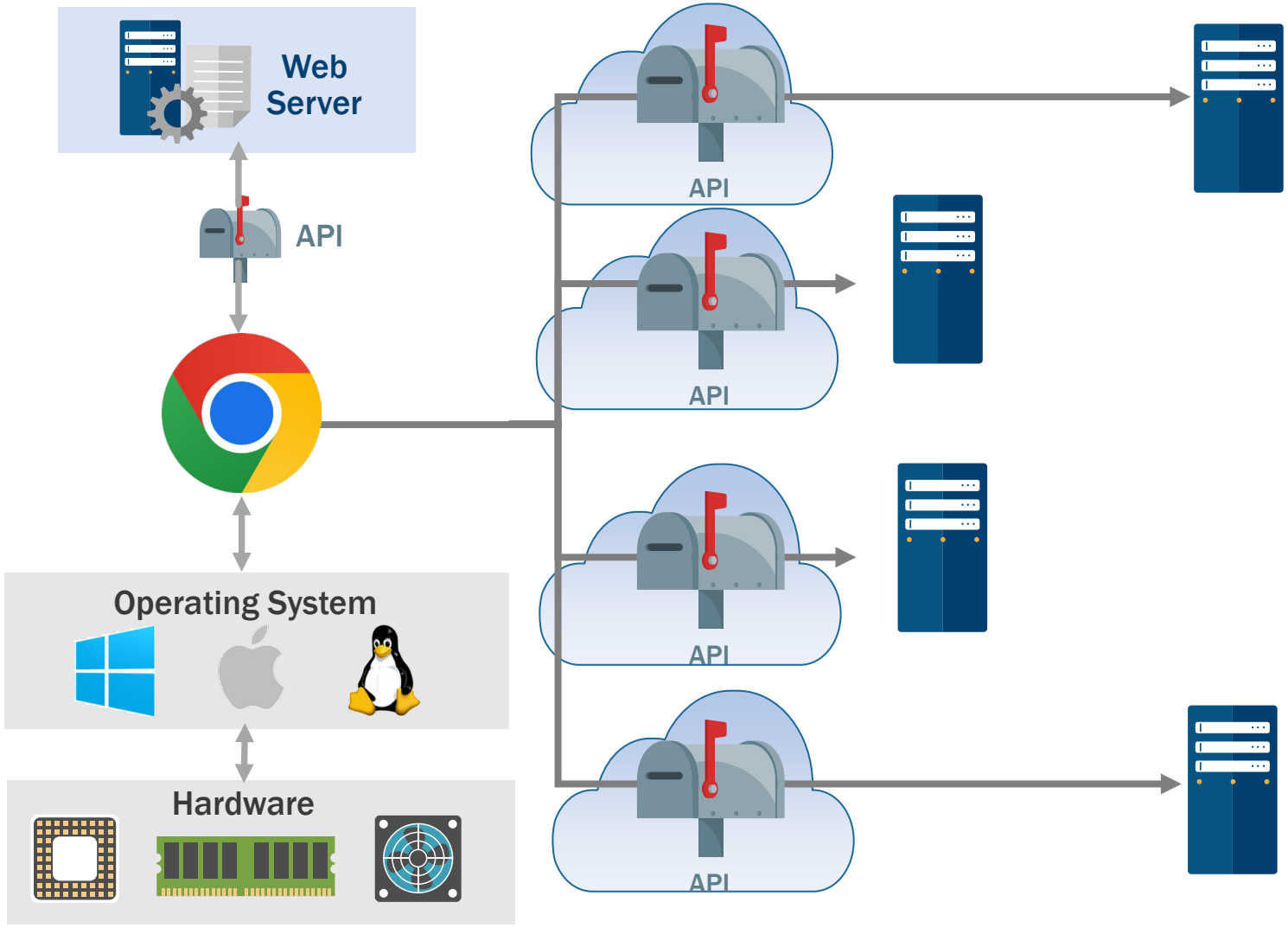


Source: PXR0321

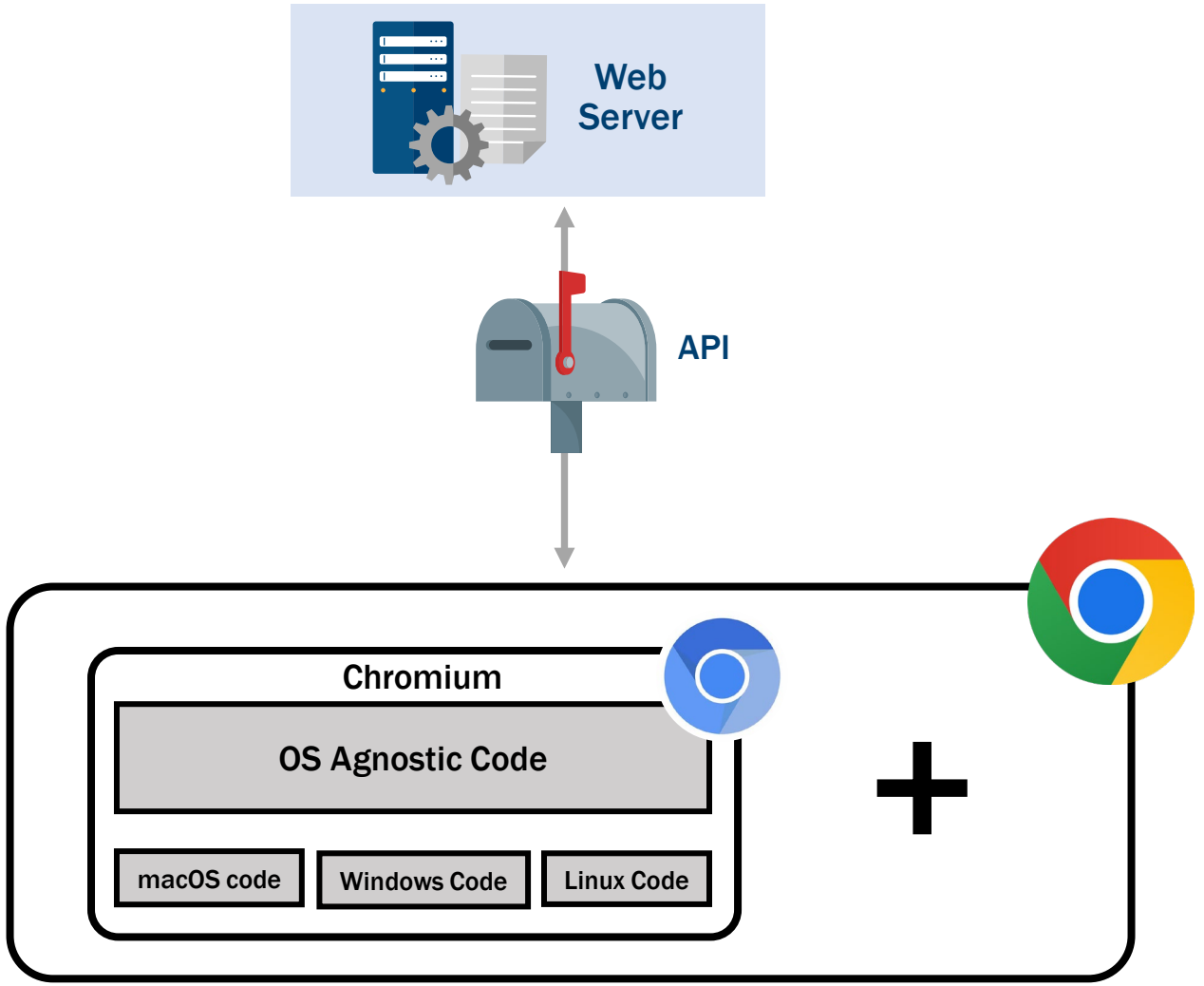
Chrome Communicates with Google Services Through Well-Defined Interfaces



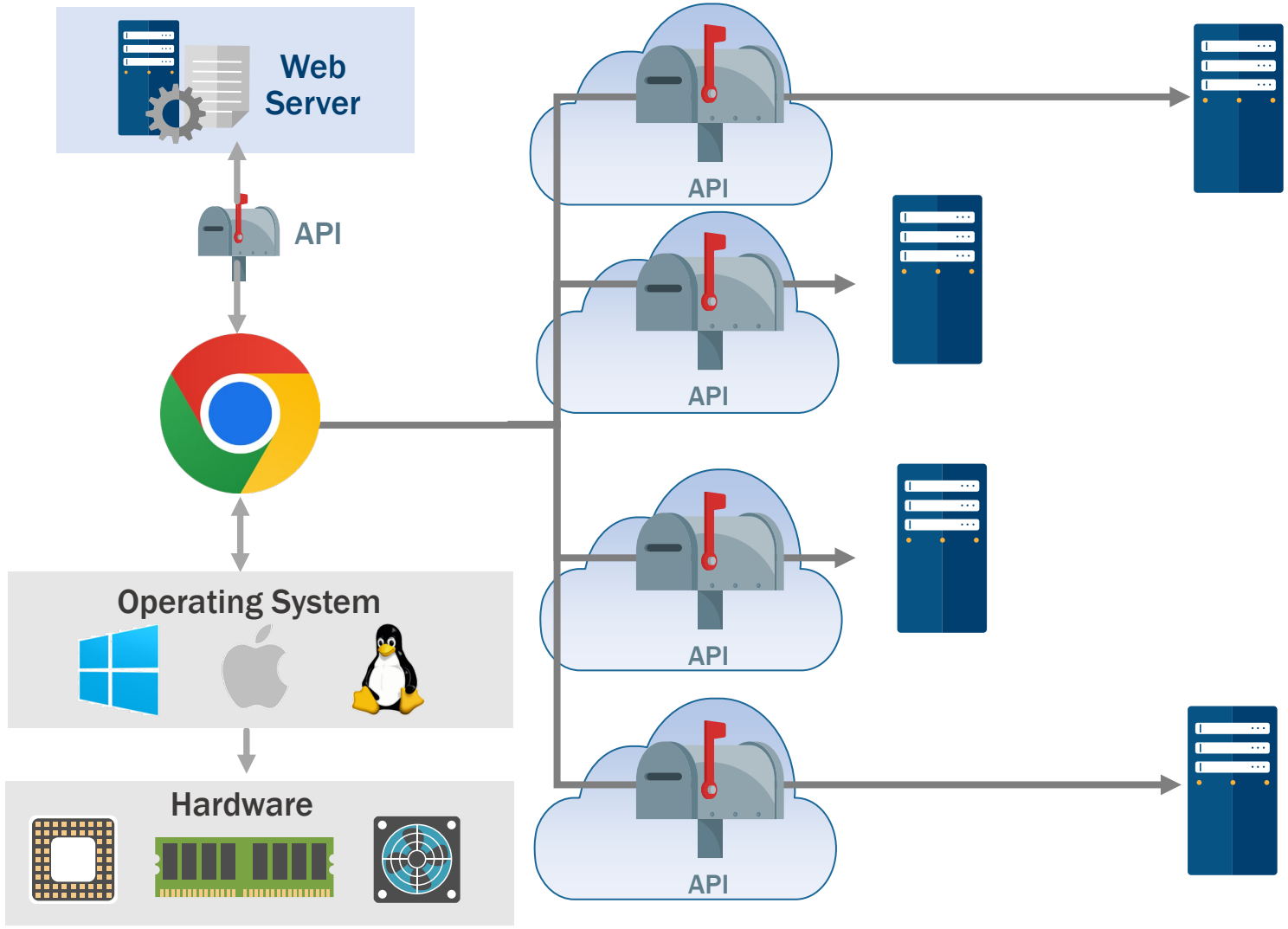
Divesting Chrome is Feasible



Divesting Chrome Would Not Impact Chrome's Ability to Retrieve and Render Web Pages



Divesting Chrome Is Feasible, Even If It May Change Chrome

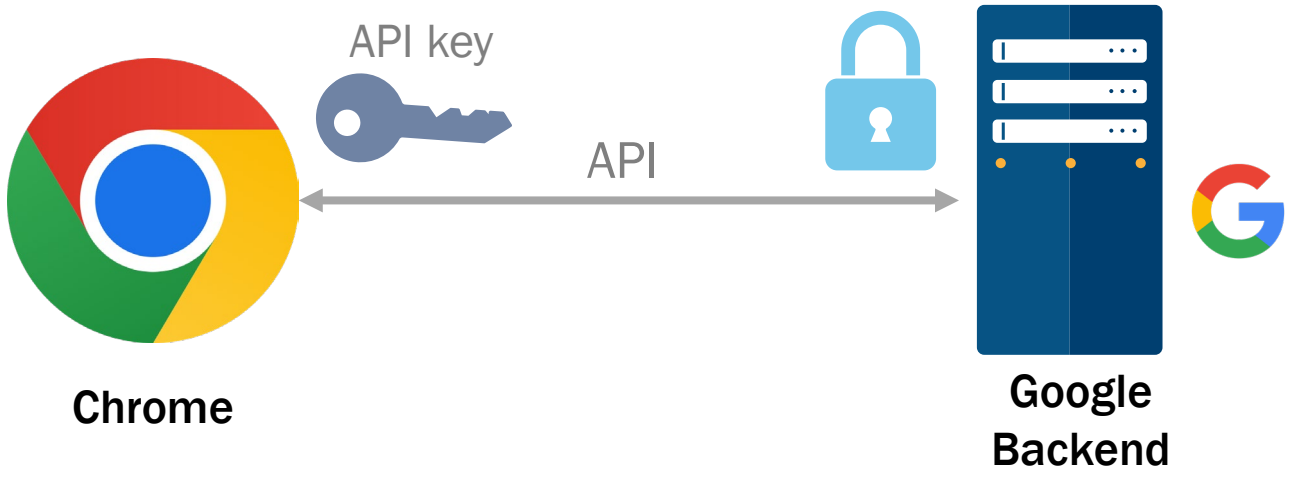


The Buyer's Four Options for Each API

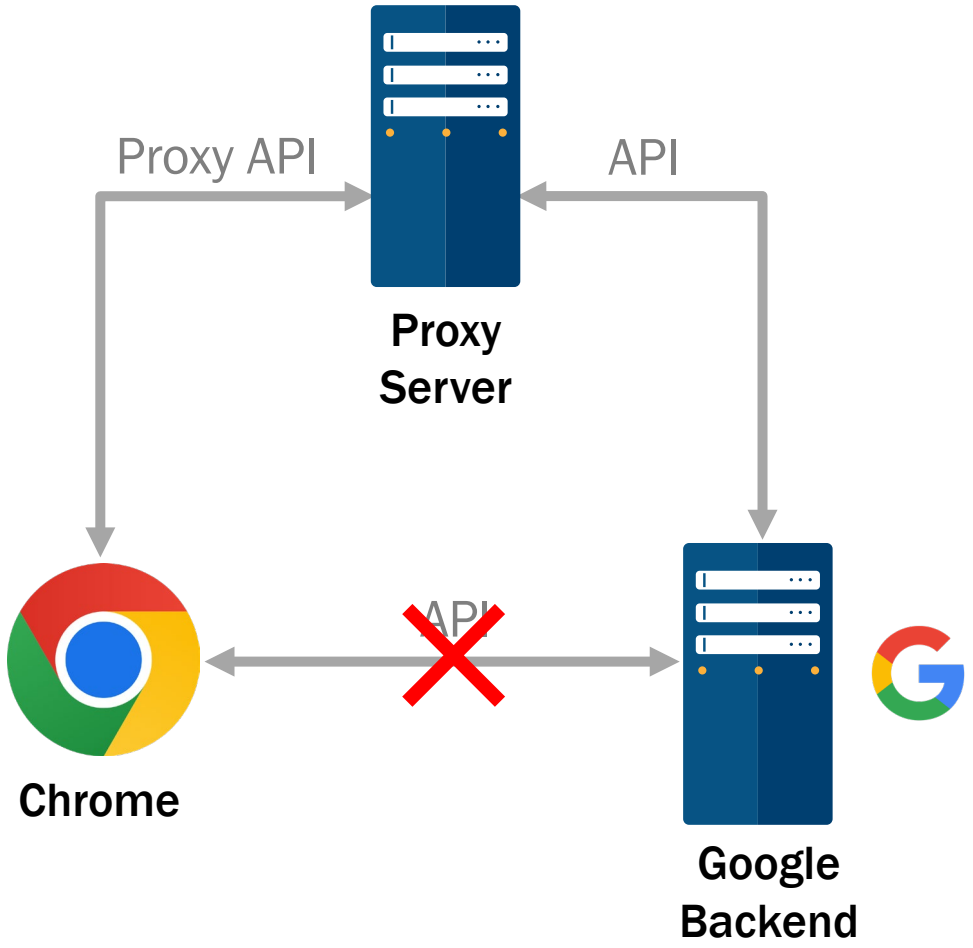
For each API, the new owner of Chrome can choose to:

- **Leave unmodified** the API call (e.g., purchase API keys)
- **Proxy** the API call to a proxy server
- **Substitute** with an API call to a server run by third-party
- **Disable** the API call

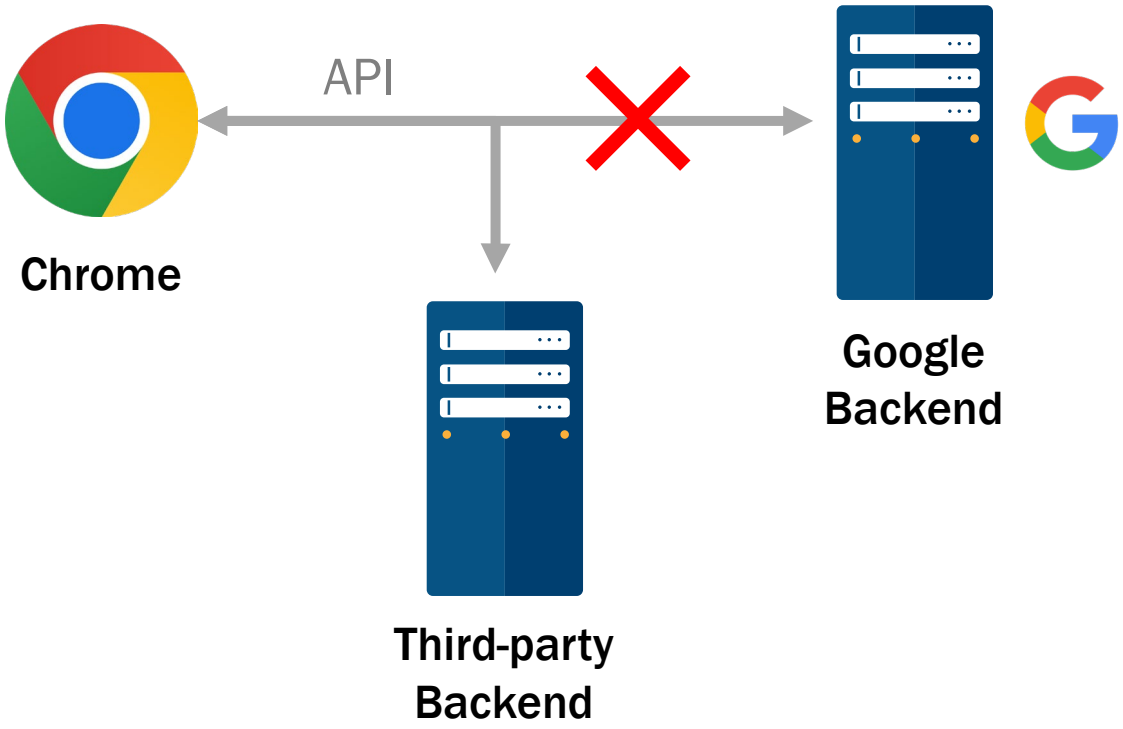
Leave Unmodified the API Call



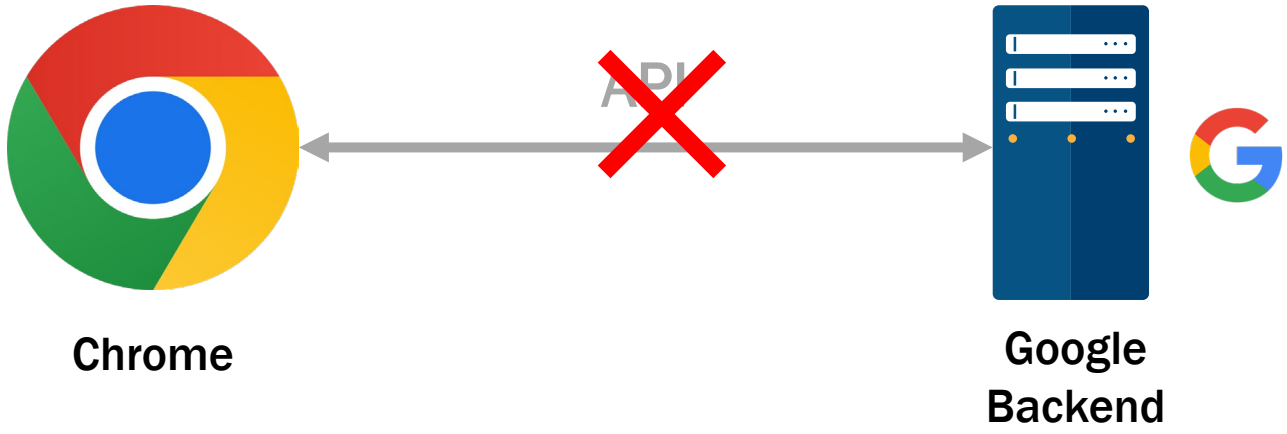
Proxy the API Call



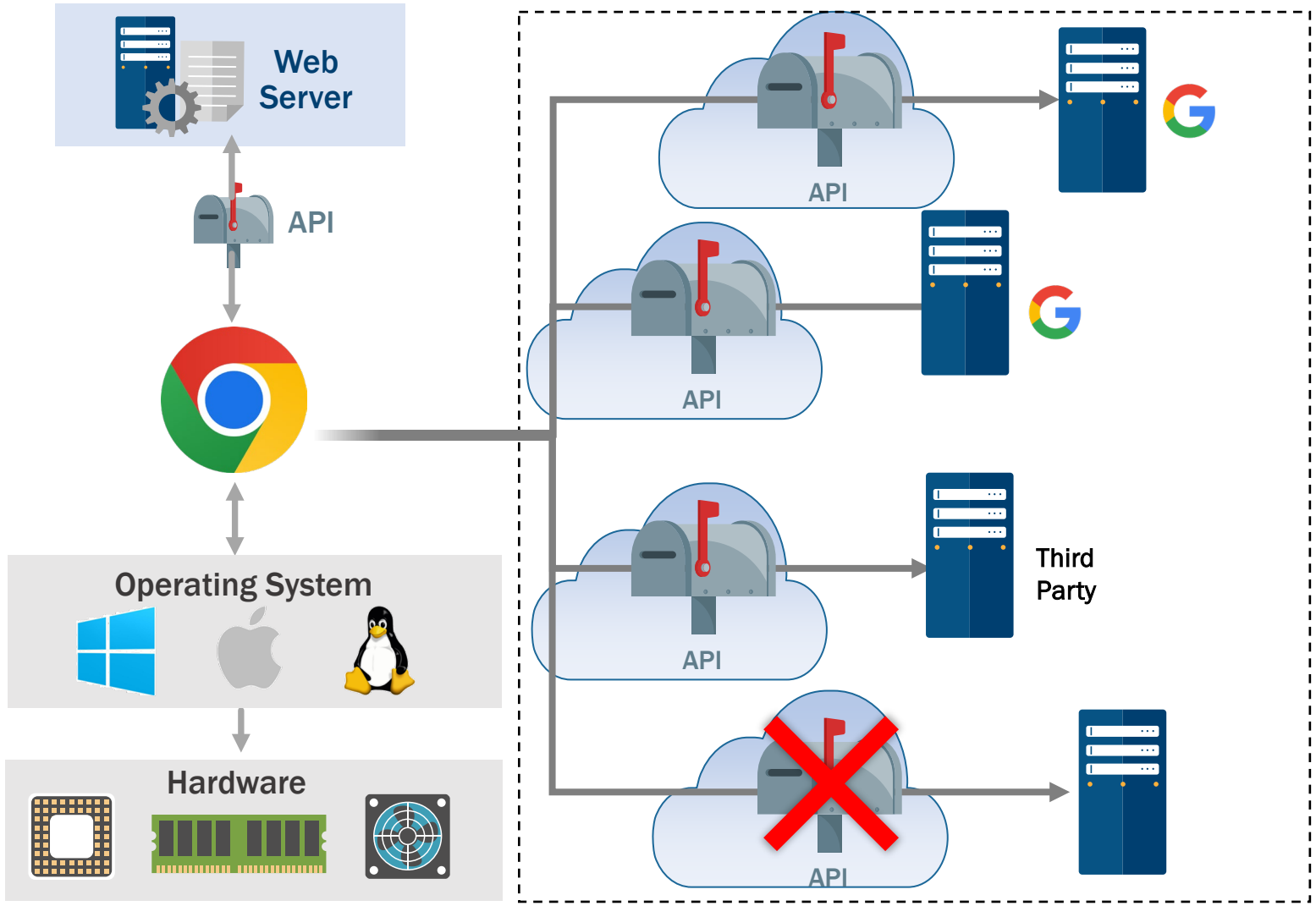
Substitute the API Call



Disable the API Call



Divesting Chrome Is Feasible, Even If It Likely Changes Chrome

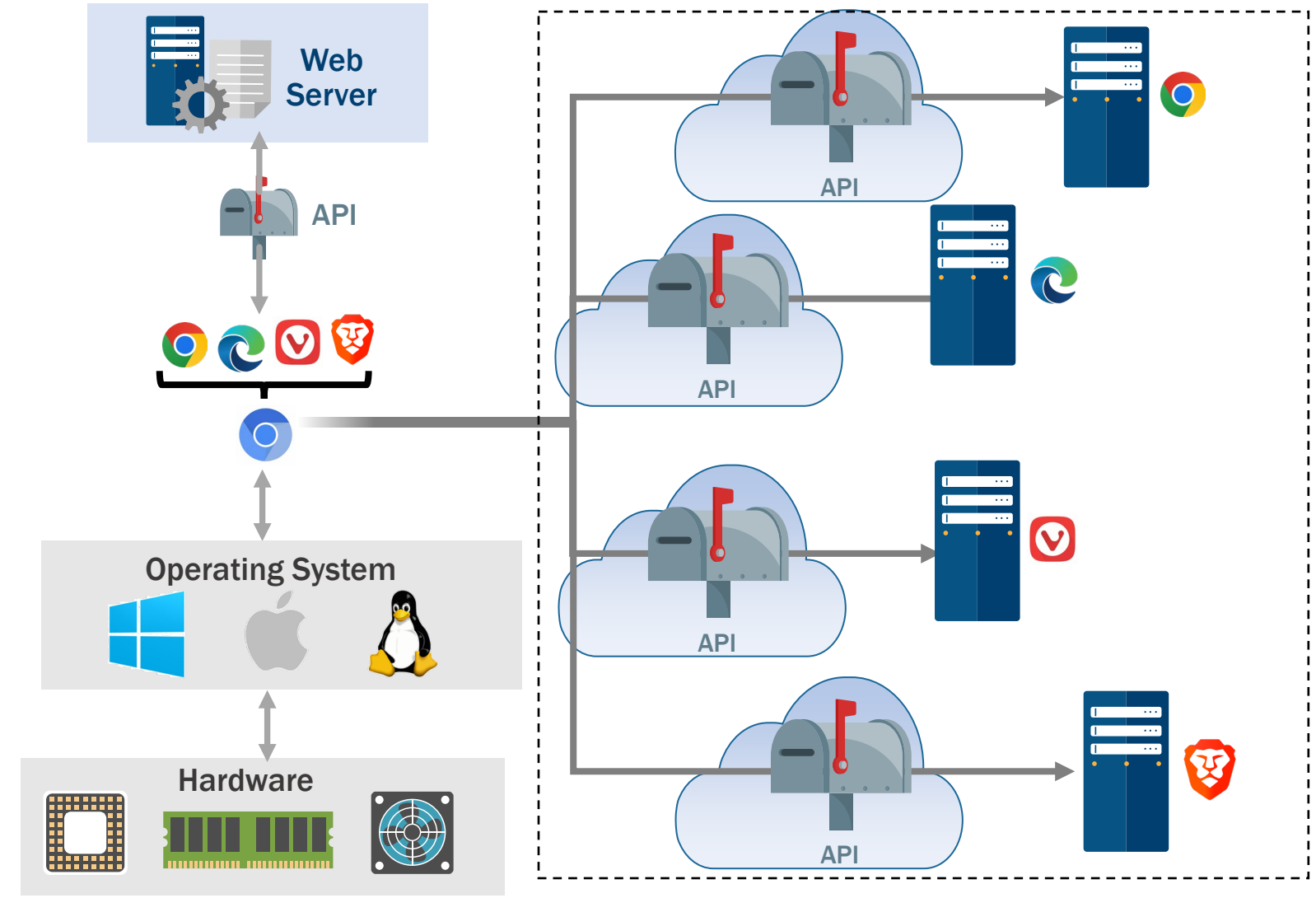


This Has All Happened Before

Services we replaced or turned off

Safe browsing	Speech input	Single sign-on (Gaia)	Developer Tools Remote Debugging
Nearby messages	Google Pay	Content Hash Fetcher	iOS Promotion Service
Link Doctor	Drive API	Flighting Service	One Google Bar Download
Ad blocking	Chrome OS hardware id	Component Updater Service	Brand Code Configuration Fetcher
User data sync	Device registration	RAPPORT service	WebRTC Logging
Spellcheck	Google Maps Time zone	Chrome OS monitor calibration	Captive Portal Service
Suggest	Google Cloud Storage	Chrome OS device management	
Translate	Cloud Print	Android app password sync	
SmartLock	Google DNS	Offline Page Service	
Form Fill	Supervised Profiles	Feedback	
Push Notifications	Address Format	Domain Reliability Monitoring	
WebStore	Network Location	Data Reduction Proxy	
Extension Store	Network Time	Chrome Cleanup	
Maps Geolocation	Favicon service		
Google Now	Google Cloud Messaging		

Microsoft



Source: Adapted from https://www.youtube.com/watch?v=zb5ggwig1aE

Post-Divestiture Chromium Will Receive Developer Attention

*"Several leading organizations have already pledged their support for the initiative, including Google, Meta, Microsoft, and Opera. **These organizations are committed to driving innovation** in the Chromium ecosystem through their involvement in this initiative."*

*"The Supporters of Chromium-Based Browsers follow an **open governance model**, drawing from best practices established by other successful Linux Foundation initiatives. It prioritizes transparency, inclusivity, and **community-driven development**."*

Source: PXR0333

This Has All Happened Before

kubernetes	A tool for managing the deployment of distributed systems. Kubernetes is an offshoot of Google's closed-source Borg project.
Firefox	A popular closed-source browser in the 1990s. Netscape created a non-profit organization called Mozilla, who then made the code open-source, rebranded it as Firefox.
blender	Blender was initially created as a closed-source program that was made open-source under the guidance of the non-profit Blender Foundation.

Source: Opening Report at 45–46, ¶ 79 (citing public facing articles).

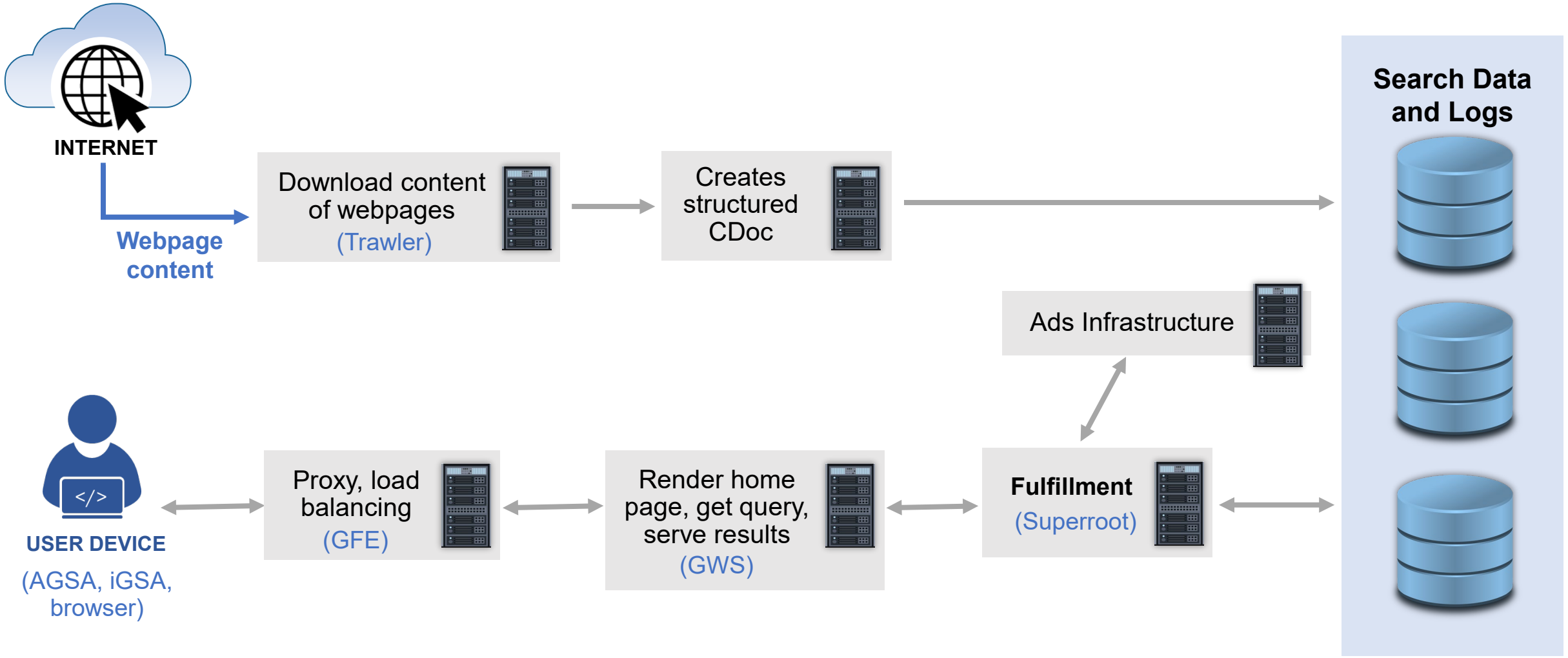
Opinion # 2: The Proposed Data Sharing and Syndication Remedies Are Feasible

Data Sharing Remedies: Google's search infrastructure is amenable to providing search engine prerequisites and ads data to Qualified Competitors.

Syndication Remedies: Google can build upon its existing search syndication systems to provide a syndicated search feed, explanatory SERP data for each syndicated query, a synthetic search feed, and a syndicated ad feed.

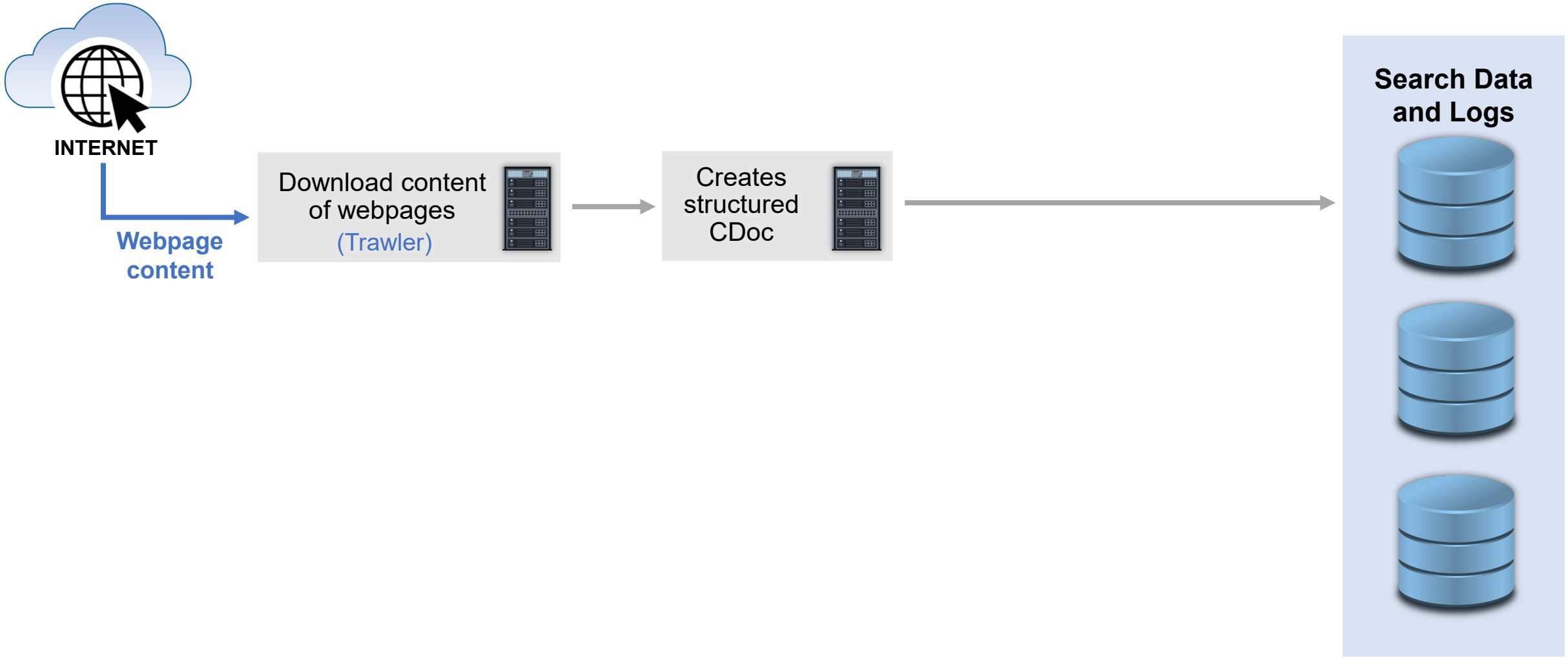
Source: Mickens Opening Report at 7–8 n.20–29 (citing RPFJ §§ VI.A,, VI.E, VI.C, VI.E, VII.A, VIII.A, VIII.C, VIII.E).

Overview of Google's Search Product

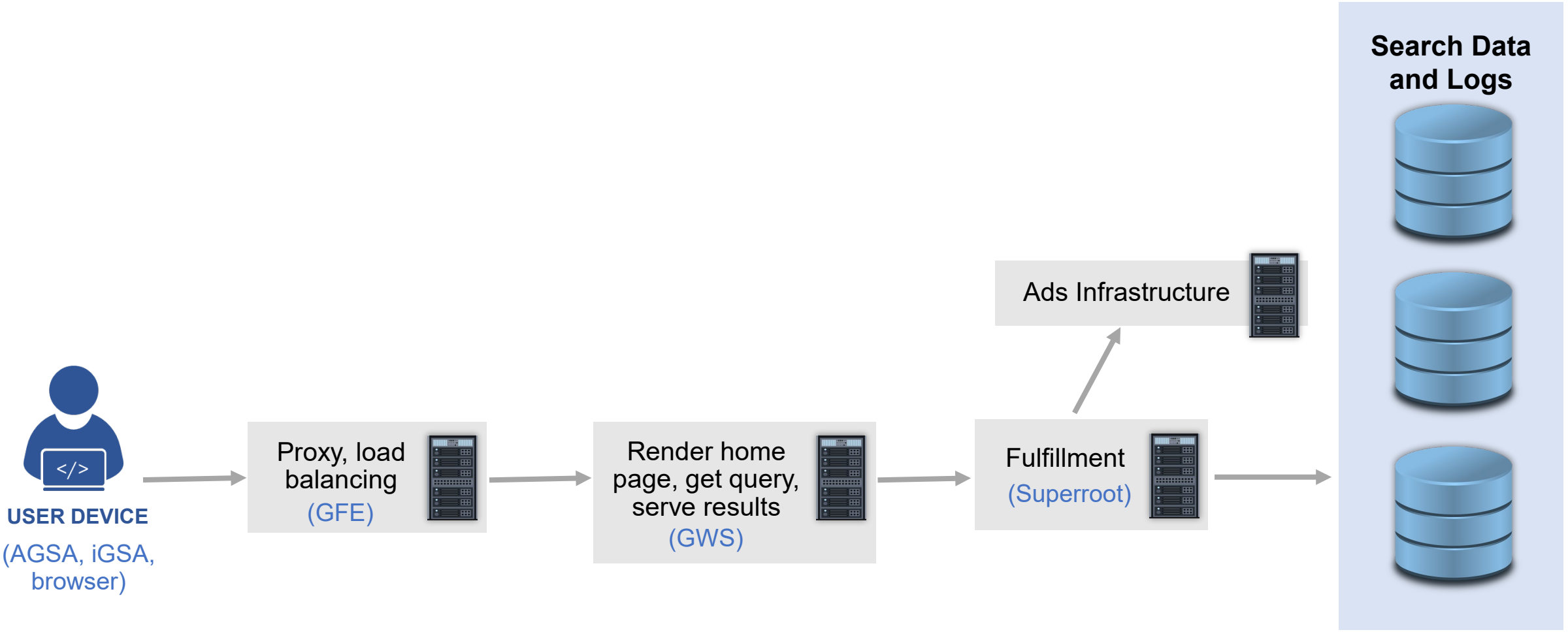


Source: Adapted from UPX0870

Indexing the Web

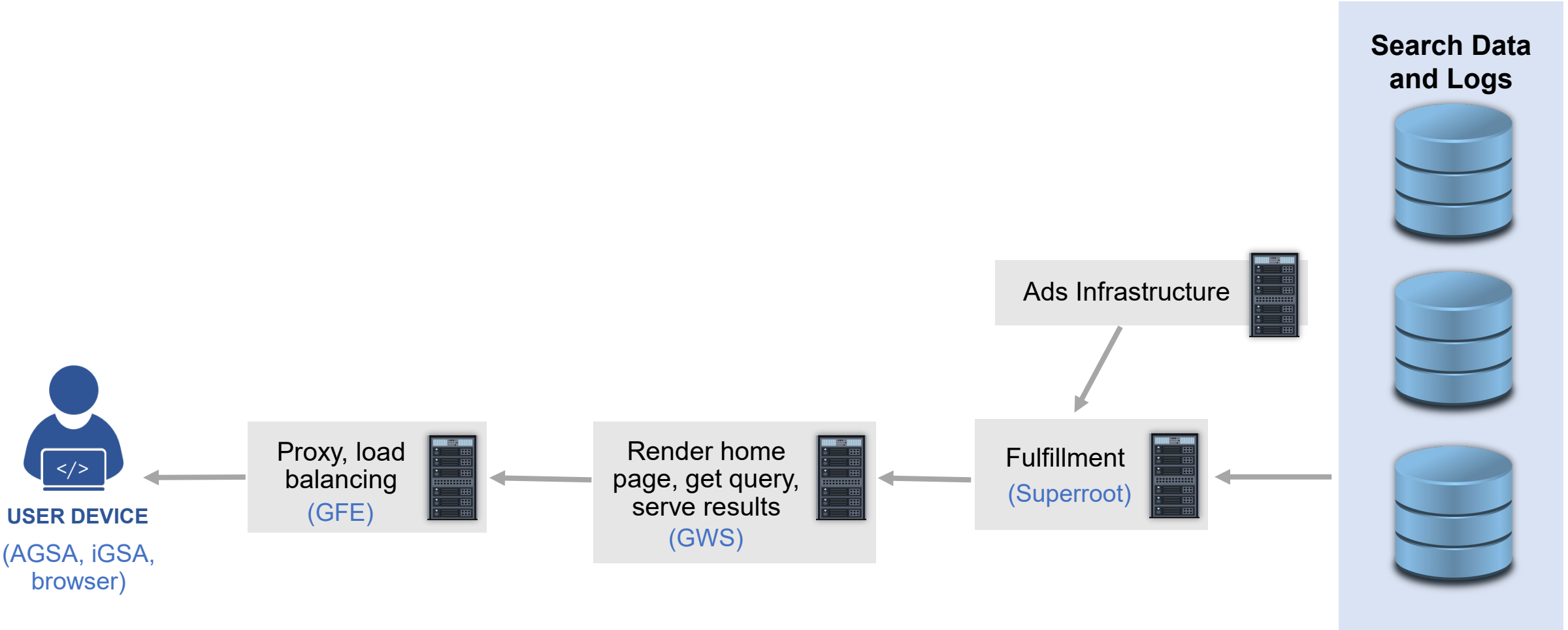


Part 1: Finding Results in Response to Queries

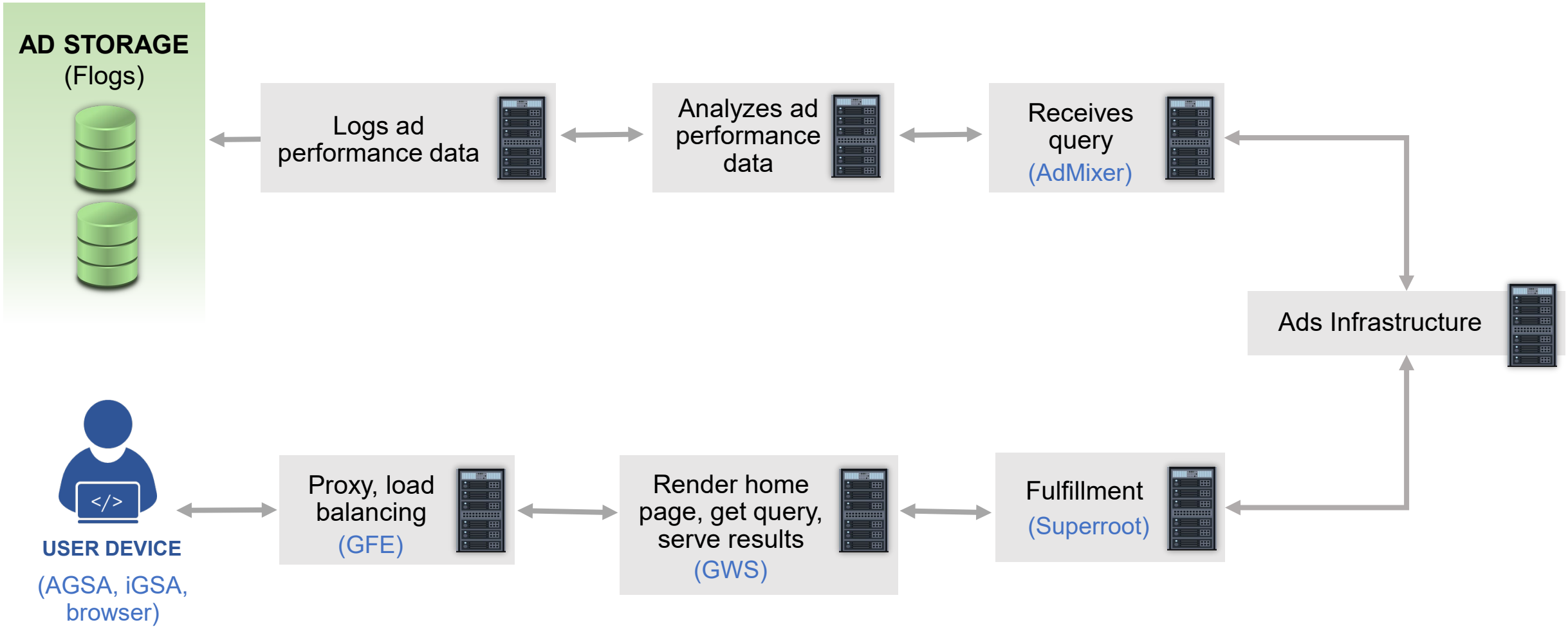


Source: Adapted from UPX0870

Part 2: Serving Search Results

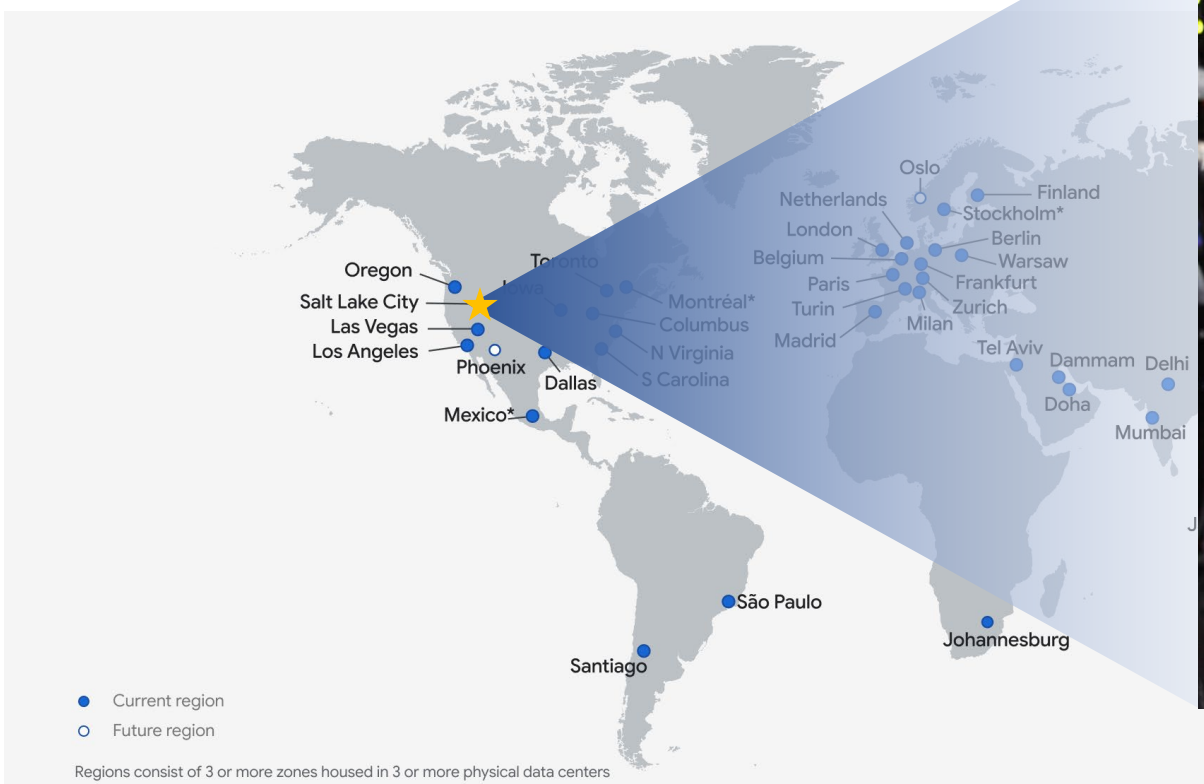


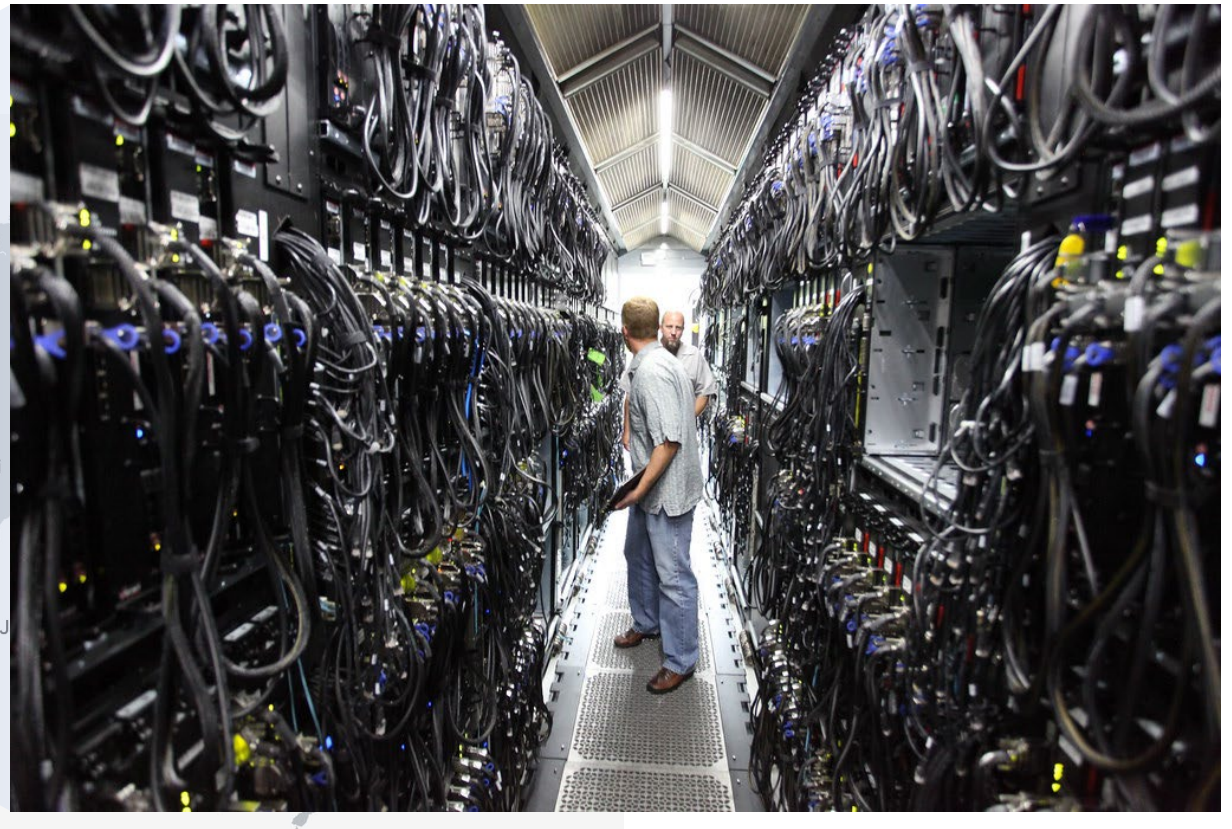
Selecting and Serving Ads



Source: Adapted from UPX0870; PXR0316 at -569

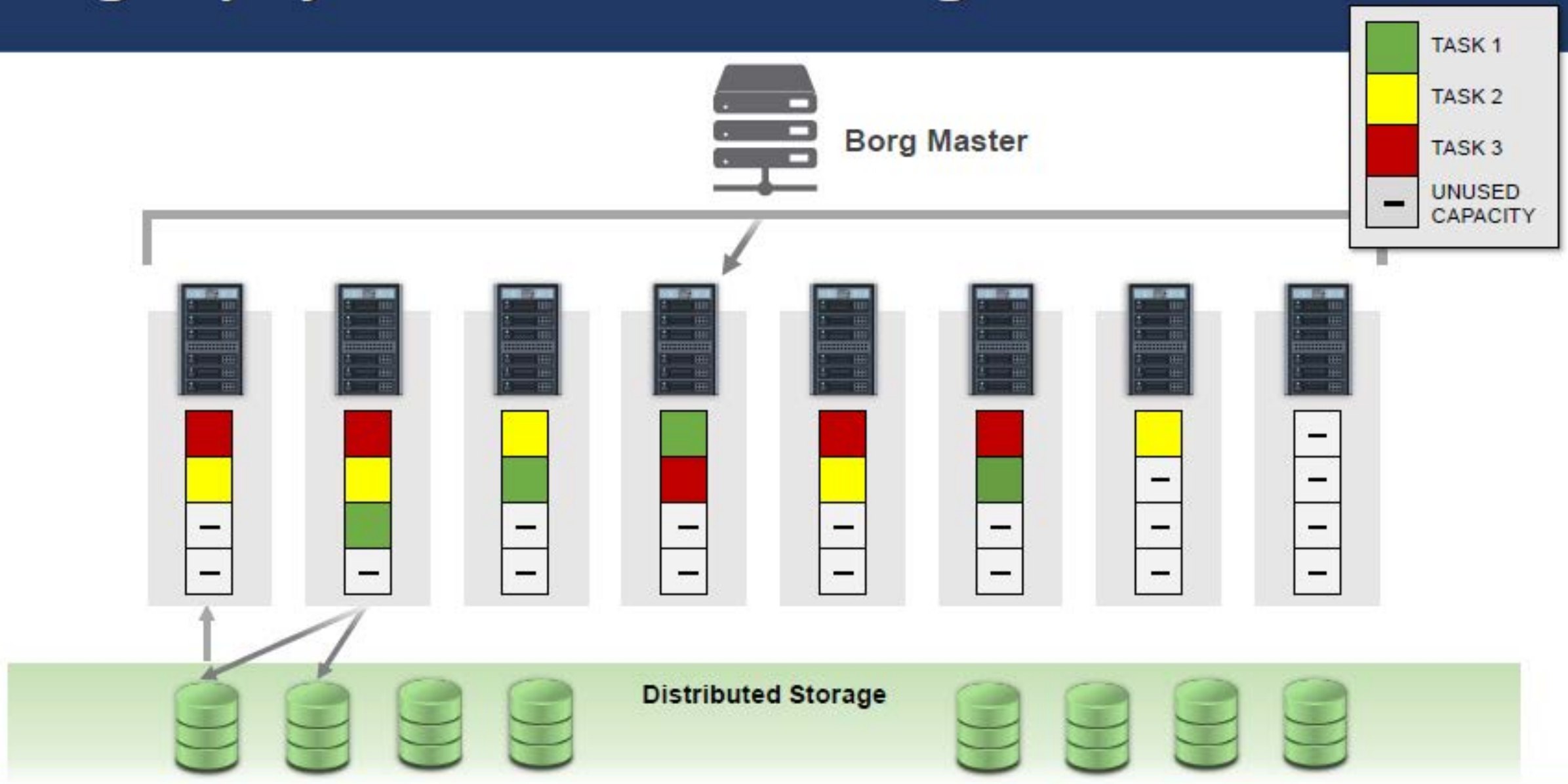
Hardware Infrastructure





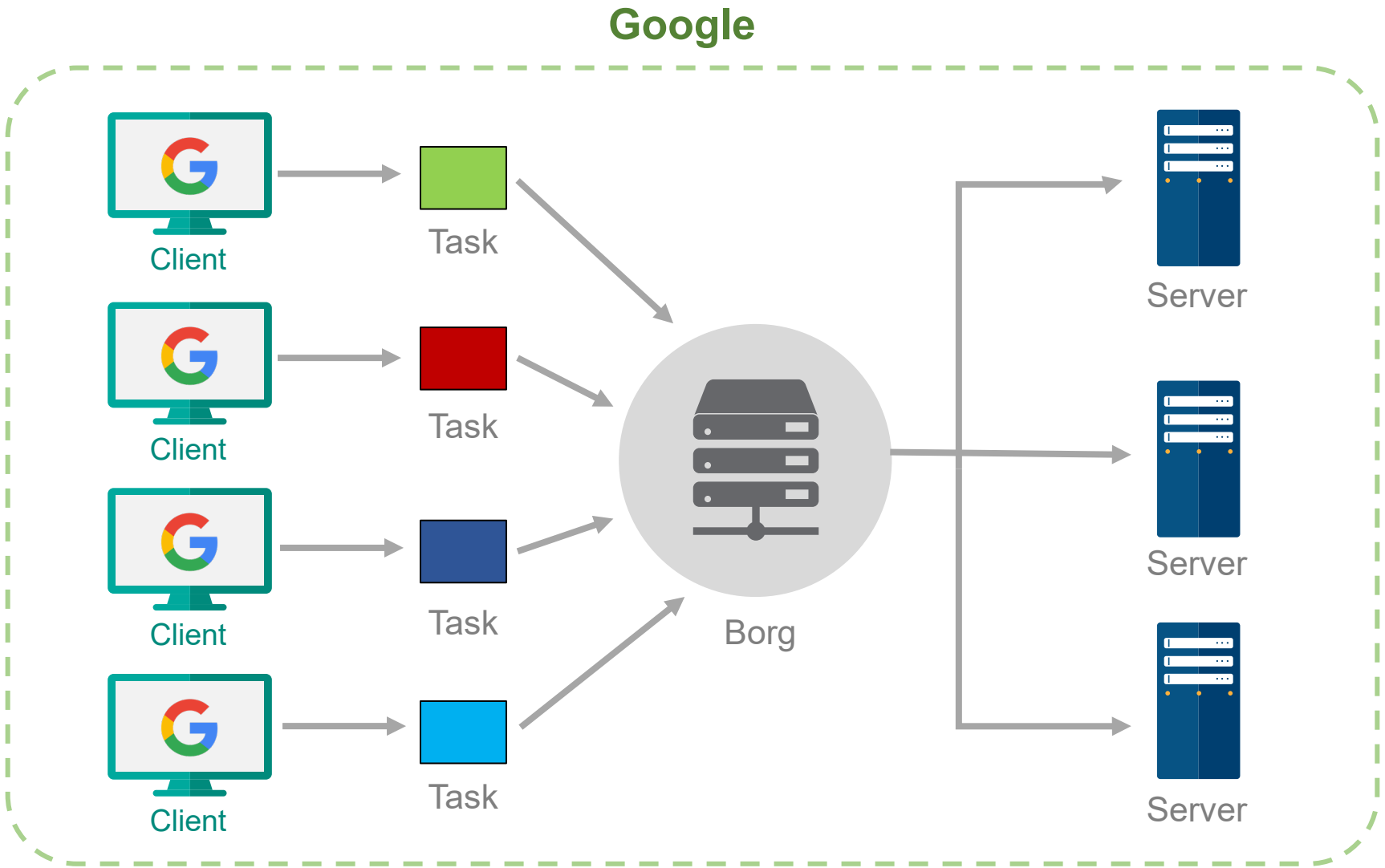
Source: https://datacenters.google/

Borg Deploys Software Across Google's Datacenters



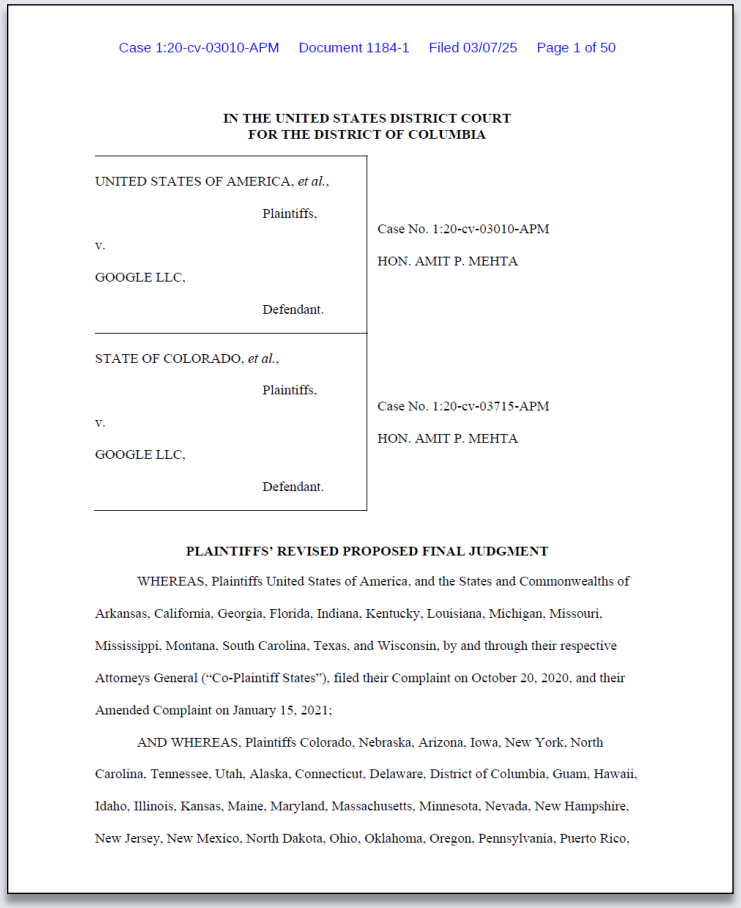
Source: Adapted from Opening Report at 69–72 (citing https://research.google/pubs/large-scale-cluster-management-at-google-with-borg/). See also

Current Borg Example



Source: Adapted from Opening Report at 69–72 (citing https://research.google/pubs/large-scale-cluster-management-at-google-with-borg/). See also

Data Sharing Remedies



Case 1:20-cv-03010-APM Document 1184-1 Filed 03/07/25 Page 1 of 50

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

UNITED STATES OF AMERICA, *et al.*, Plaintiffs, v. GOOGLE LLC, Defendant.	Case No. 1:20-cv-03010-APM HON. AMIT P. MEHTA
STATE OF COLORADO, *et al.*, Plaintiffs, v. GOOGLE LLC, Defendant.	Case No. 1:20-cv-03715-APM HON. AMIT P. MEHTA

PLAINTIFFS' REVISED PROPOSED FINAL JUDGMENT

WHEREAS, Plaintiffs United States of America, and the States and Commonwealths of Arkansas, California, Georgia, Florida, Indiana, Kentucky, Louisiana, Michigan, Missouri, Mississippi, Montana, South Carolina, Texas, and Wisconsin, by and through their respective Attorneys General ("Co-Plaintiff States"), filed their Complaint on October 20, 2020, and their Amended Complaint on January 15, 2021;

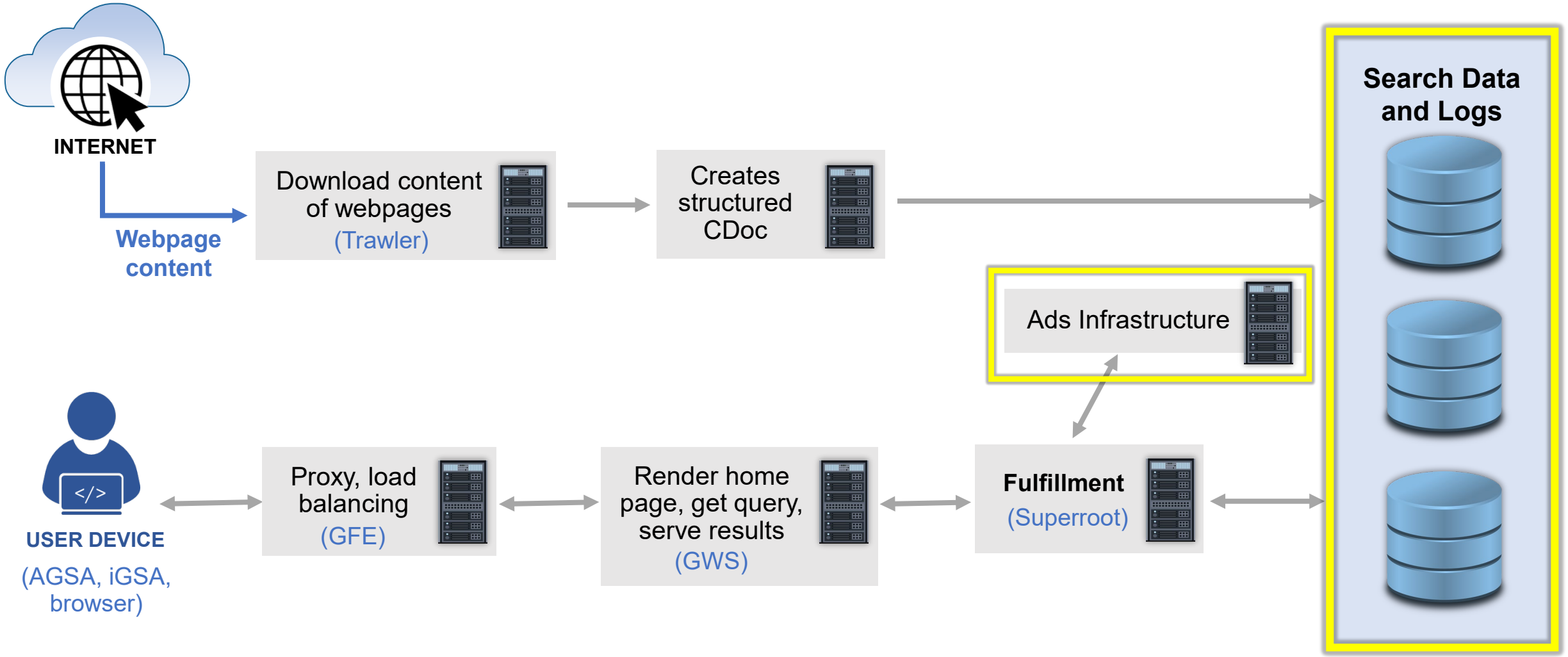
AND WHEREAS, Plaintiffs Colorado, Nebraska, Arizona, Iowa, New York, North Carolina, Tennessee, Utah, Alaska, Connecticut, Delaware, District of Columbia, Guam, Hawaii, Idaho, Illinois, Kansas, Maine, Maryland, Massachusetts, Minnesota, Nevada, New Hampshire, New Jersey, New Mexico, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Puerto Rico,

VI. REQUIRED DISCLOSURES OF SCALE-DEPENDENT DATA NECESSARY TO COMPETE WITH GOOGLE

VIII. SEARCH TEXT AD TRANSPARENCY AND REDUCTION OF SWITCHING COSTS

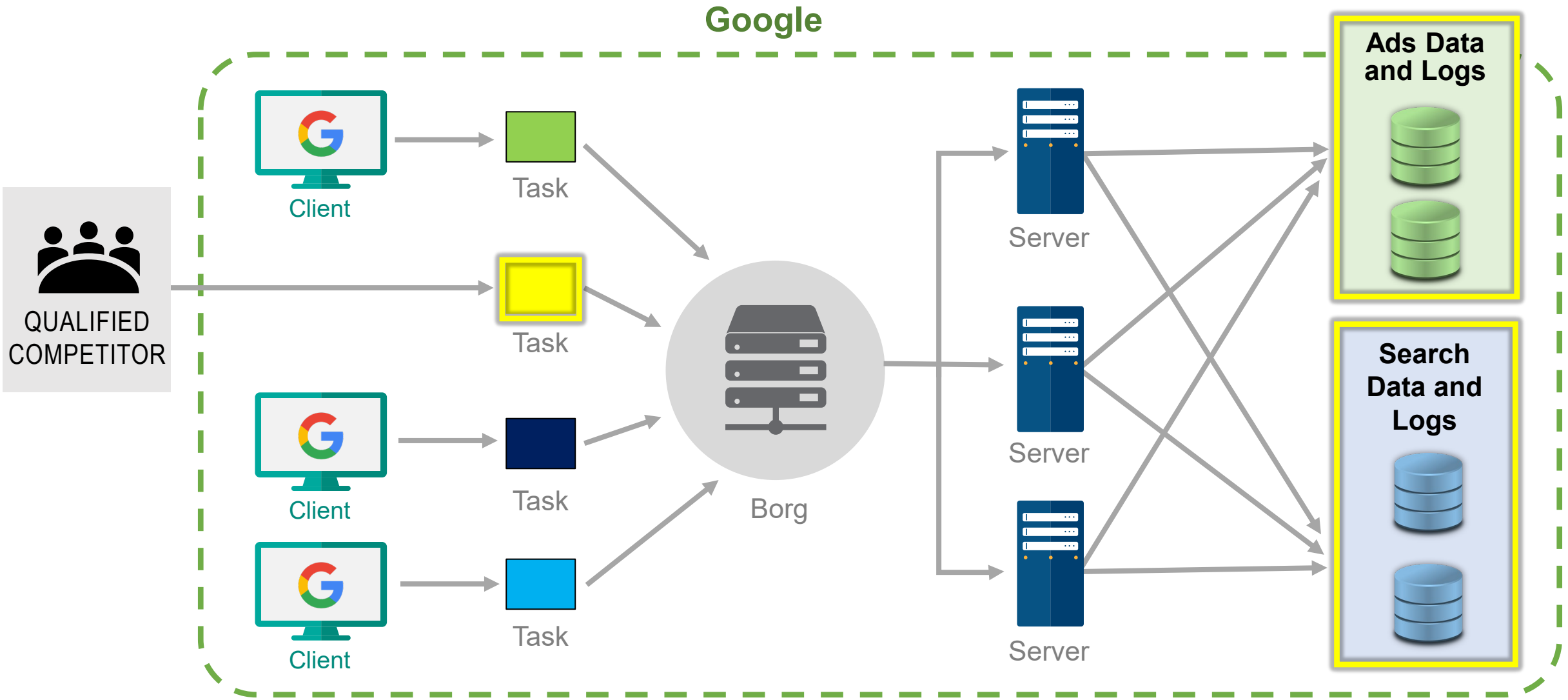
Source: Plaintiffs' RPFJ §§ VI, VIII

Data Sharing Remedies



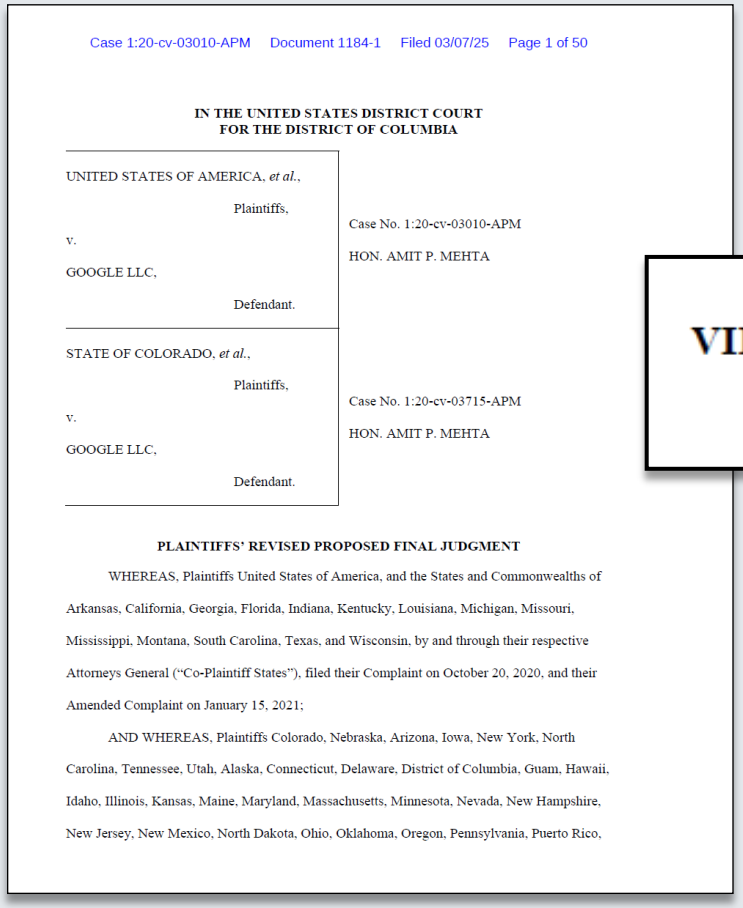
Source: Adapted from UPX0870; PXR0316 at -569

Google's Existing Infrastructure Supports Data Sharing Remedies



Source: Adapted from UPX0870; PXR0316 at -569

Data Syndication Remedies



Case 1:20-cv-03010-APM Document 1184-1 Filed 03/07/25 Page 1 of 50

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

UNITED STATES OF AMERICA, *et al.*,

Plaintiffs,

v.

GOOGLE LLC,

Defendant.

Case No. 1:20-cv-03010-APM

HON. AMIT P. MEHTA

STATE OF COLORADO, *et al.*,

Plaintiffs,

v.

GOOGLE LLC,

Defendant.

Case No. 1:20-cv-03715-APM

HON. AMIT P. MEHTA

PLAINTIFFS' REVISED PROPOSED FINAL JUDGMENT

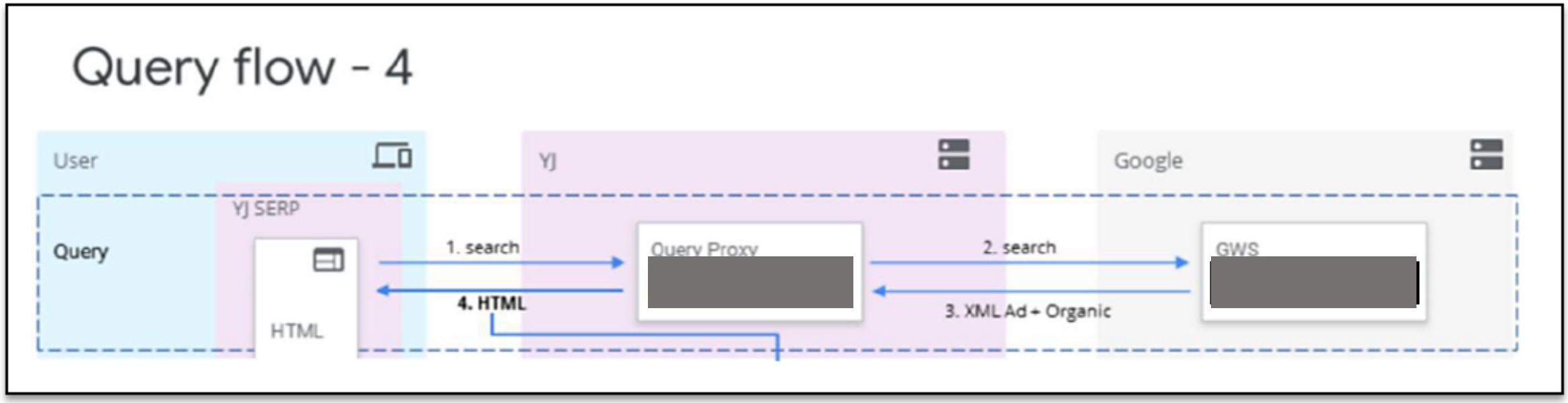
WHEREAS, Plaintiffs United States of America, and the States and Commonwealths of Arkansas, California, Georgia, Florida, Indiana, Kentucky, Louisiana, Michigan, Missouri, Mississippi, Montana, South Carolina, Texas, and Wisconsin, by and through their respective Attorneys General ("Co-Plaintiff States"), filed their Complaint on October 20, 2020, and their Amended Complaint on January 15, 2021;

AND WHEREAS, Plaintiffs Colorado, Nebraska, Arizona, Iowa, New York, North Carolina, Tennessee, Utah, Alaska, Connecticut, Delaware, District of Columbia, Guam, Hawaii, Idaho, Illinois, Kansas, Maine, Maryland, Massachusetts, Minnesota, Nevada, New Hampshire, New Jersey, New Mexico, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Puerto Rico,

VII. REQUIRED SYNDICATION OF SEARCH RESULTS NECESSARY TO BUILD GSE QUALITY AND SCALE OF QUALIFIED COMPETITORS

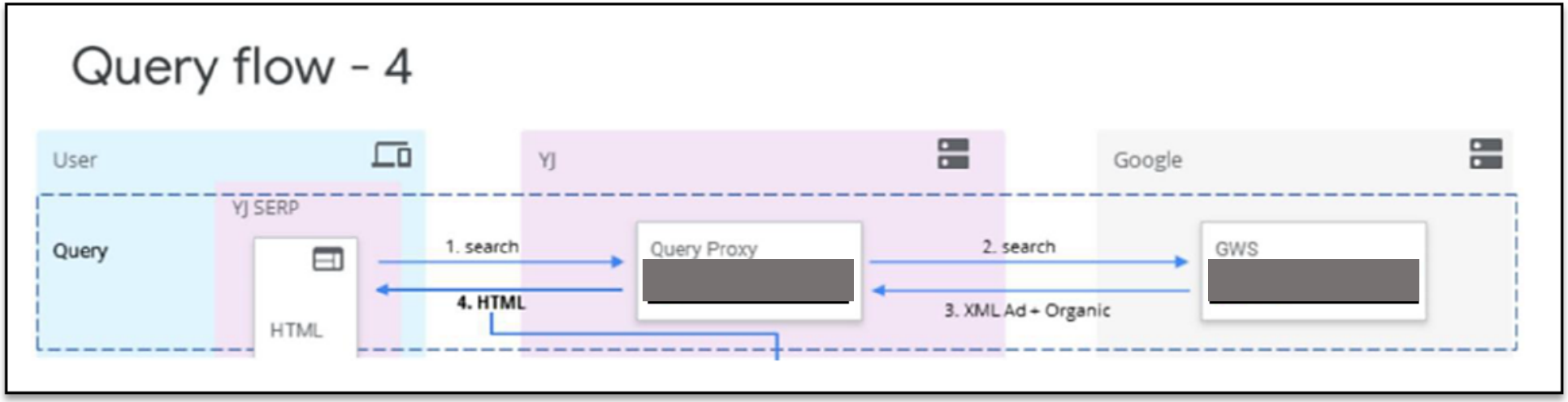
Source: Plaintiffs' RPFJ § VII

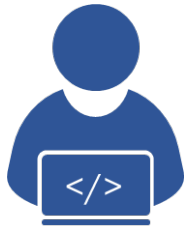
This Has All Happened Before



Source: PXR0318, at -083

There Is Technical Precedent for Syndication





QC USER DEVICE
(AGSA, iGSA, Browser)

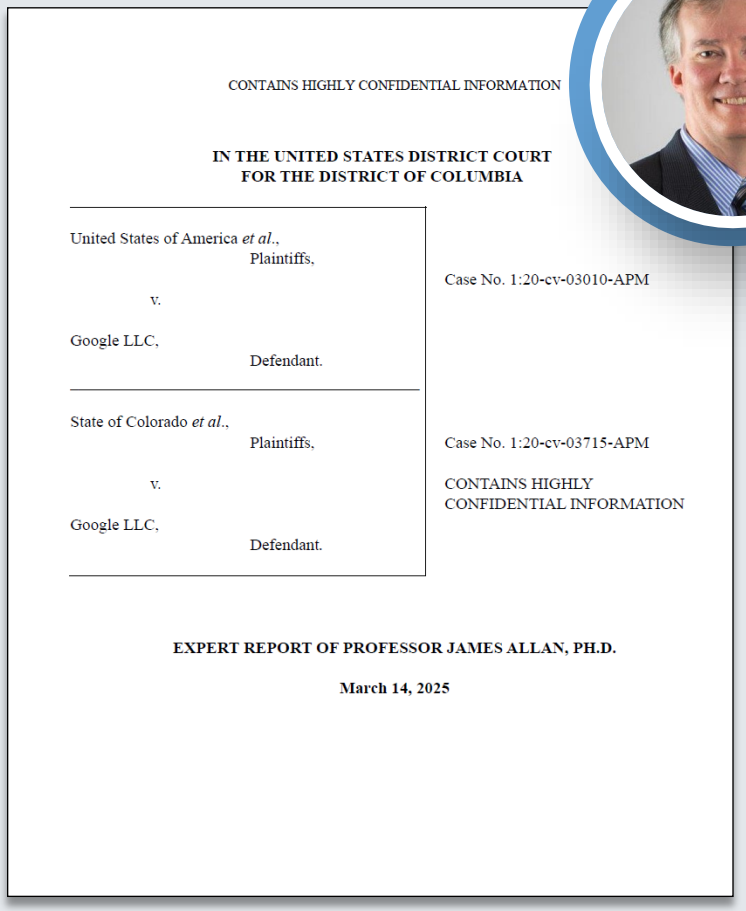
QC Infra

Google

API

In response to query
(GWS)

Professor Allan's Claims on Reverse-Engineering Are Misleading



CONTAINS HIGHLY CONFIDENTIAL INFORMATION

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

United States of America *et al.*,
Plaintiffs,

v.

Google LLC,
Defendant.

Case No. 1:20-cv-03010-APM

State of Colorado *et al.*,
Plaintiffs,

v.

Google LLC,
Defendant.

Case No. 1:20-cv-03715-APM

CONTAINS HIGHLY CONFIDENTIAL INFORMATION

EXPERT REPORT OF PROFESSOR JAMES ALLAN, PH.D.

March 14, 2025

18. Given the breadth of Plaintiffs' proposals and recent advances in LLMs, implementing Plaintiffs' proposals would allow a competitor to reverse engineer essentially all of Google's search results or, at a minimum, reverse engineer Google's current and future technologies (Section IX). In particular, I find that:

Source: Prof. Allan Initial Report at ¶ 18

Professor Allan's Claims on Reverse-Engineering Are Misleading

1. QCs *do not want* to build a perfect replica of Google.

- OpenAI and DuckDuckGo representatives have testified they *do not want* to replicate google.com.

2. A QC *could not* solely rely on syndicated feeds and logs to replicate a search engine.

- **People**: engineers are needed to write code to replicate a service
- **Time**: engineers must design, build, deploy, and test code
- **Infrastructure**: code needs to run on physical datacenter infrastructure

Thus, QCs would be in a perpetual game of catch-up.

Professor Allan's Deposition Testimony

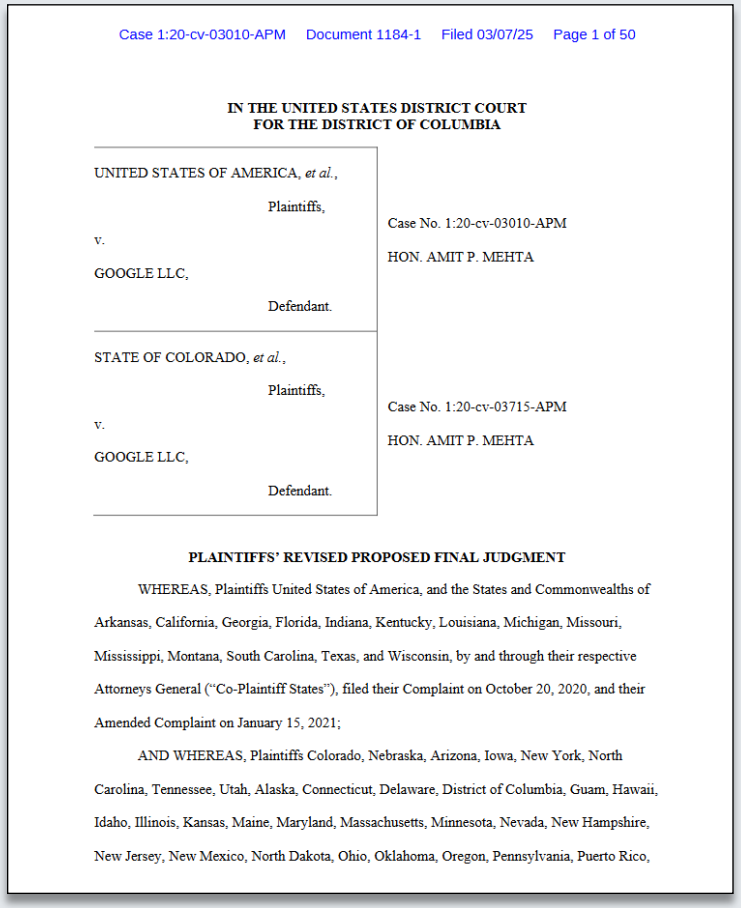
Q. So you're not offering an opinion on quality infrastructure latency that it would take a qualified competitor using this type of LLM or actually any -- any of the reverse engineering methodologies in your report?

A. Right. All of what I'm focusing on is what is disclosed, whether directly or reverse engineering or mimicking, and what a competitor -- you know, and that these are -- these are complicated technologies that are being disclosed.

And I'm not saying – I'm saying the competitor could use that to improve their system. I'm not saying whether their system would then be fast, and I have not investigated how fast. Because there's too many variables in that. It's too much of a I don't know what hardware they might have. I don't know what sort of engineering expertise they have inhouse. It's just -- there are too many variables there to offer that opinion.

Source: Allan Dep. at 261:24-263:2

Prohibition on Self-Preferencing in the RPFJ



Case 1:20-cv-03010-APM Document 1184-1 Filed 03/07/25 Page 1 of 50

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

UNITED STATES OF AMERICA, *et al.*,
Plaintiffs,
v.
GOOGLE LLC,
Defendant.

Case No. 1:20-cv-03010-APM
HON. AMIT P. MEHTA

STATE OF COLORADO, *et al.*,
Plaintiffs,
v.
GOOGLE LLC,
Defendant.

Case No. 1:20-cv-03715-APM
HON. AMIT P. MEHTA

PLAINTIFFS' REVISED PROPOSED FINAL JUDGMENT

WHEREAS, Plaintiffs United States of America, and the States and Commonwealths of Arkansas, California, Georgia, Florida, Indiana, Kentucky, Louisiana, Michigan, Missouri, Mississippi, Montana, South Carolina, Texas, and Wisconsin, by and through their respective Attorneys General ("Co-Plaintiff States"), filed their Complaint on October 20, 2020, and their Amended Complaint on January 15, 2021;

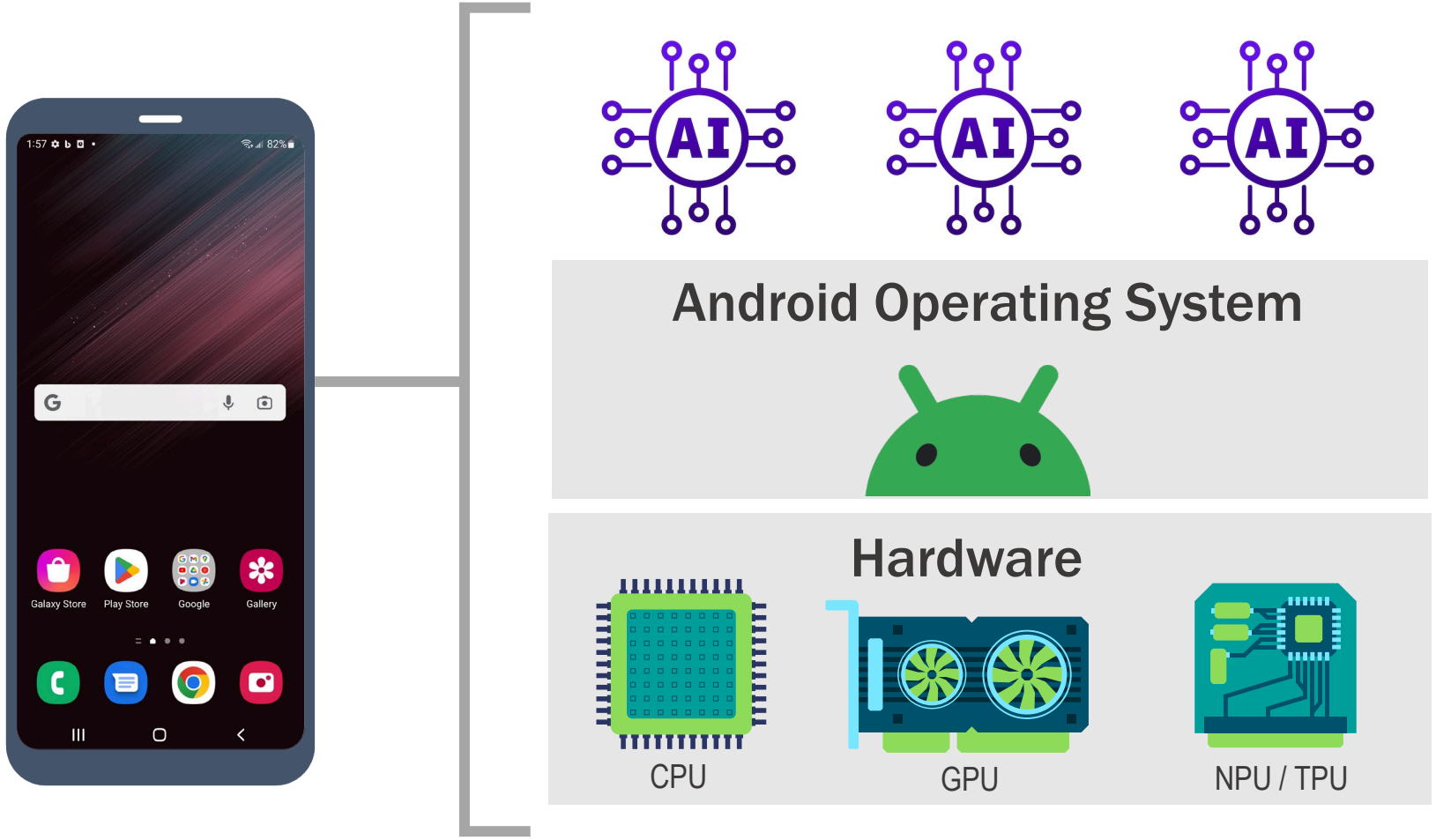
AND WHEREAS, Plaintiffs Colorado, Nebraska, Arizona, Iowa, New York, North Carolina, Tennessee, Utah, Alaska, Connecticut, Delaware, District of Columbia, Guam, Hawaii, Idaho, Illinois, Kansas, Maine, Maryland, Massachusetts, Minnesota, Nevada, New Hampshire, New Jersey, New Mexico, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Puerto Rico,

B. Self-Preferencing Prohibited: Google must not use its ownership and control of Android, or any other Google product or service, to:

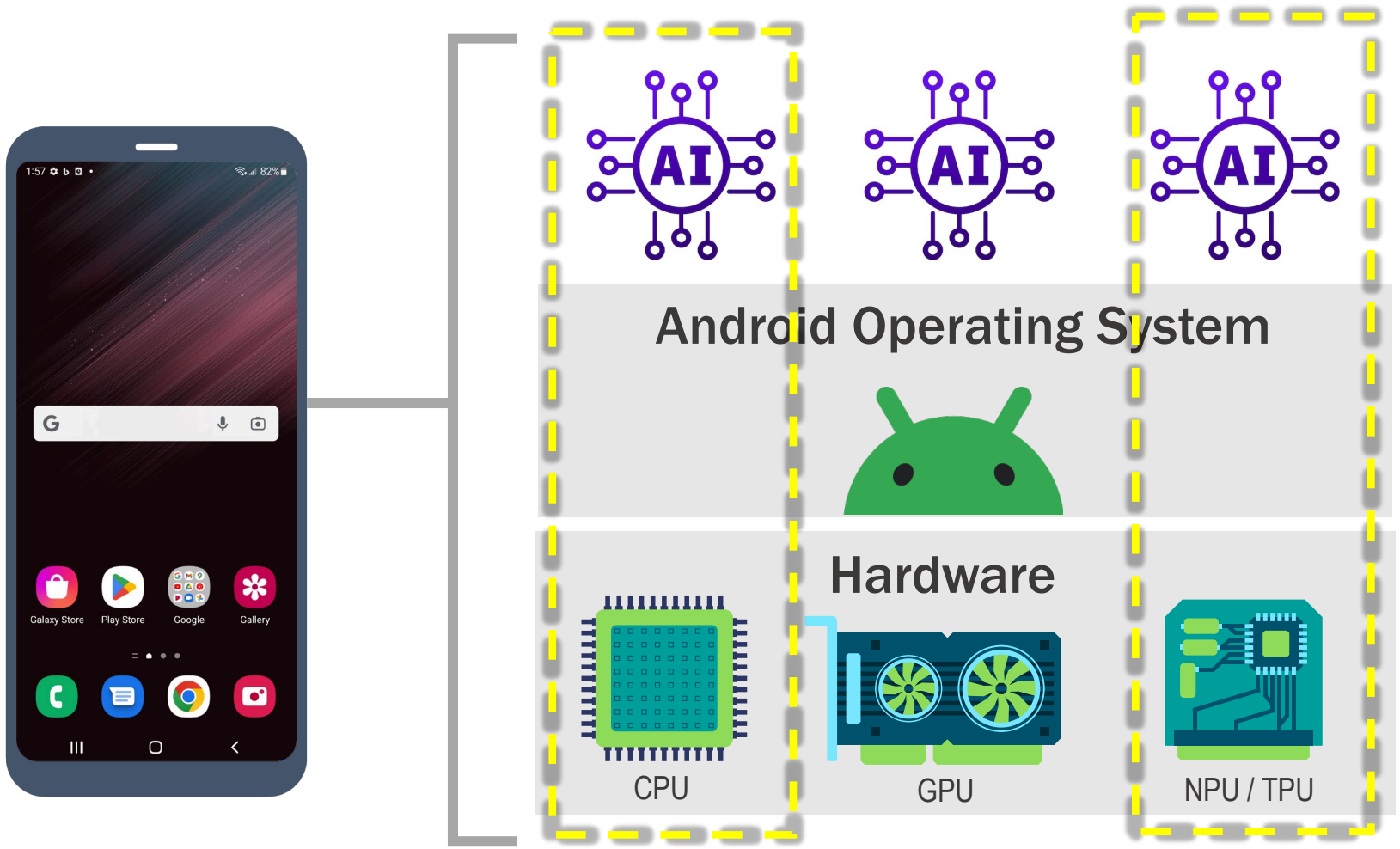
1. make any GSE, Search Access Point, GenAI Product, or On-Device AI explicitly or implicitly mandatory on Android Devices, for example, by preventing interoperability between Android AICore or a Google Grounding API and Competitor products and services in the GSE or Search Text Ads markets;

Source: Plaintiffs' RPFJ § V.B

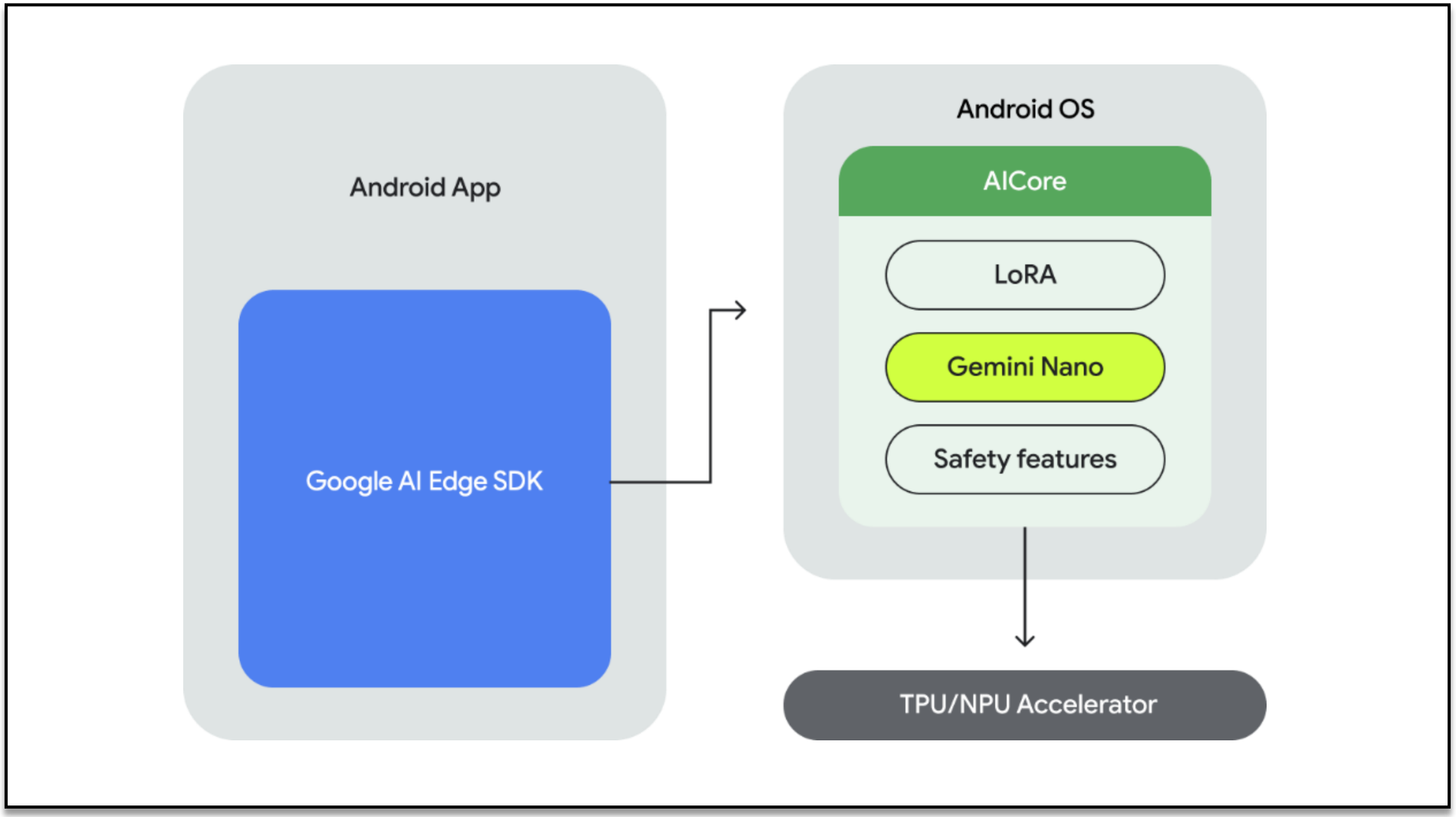
Multiple AI Models Can Run on an Android Phone



AI Models Run More Efficiently If They Run on Specialized Hardware



What Is AICore?



Source: PXR0310 at -453

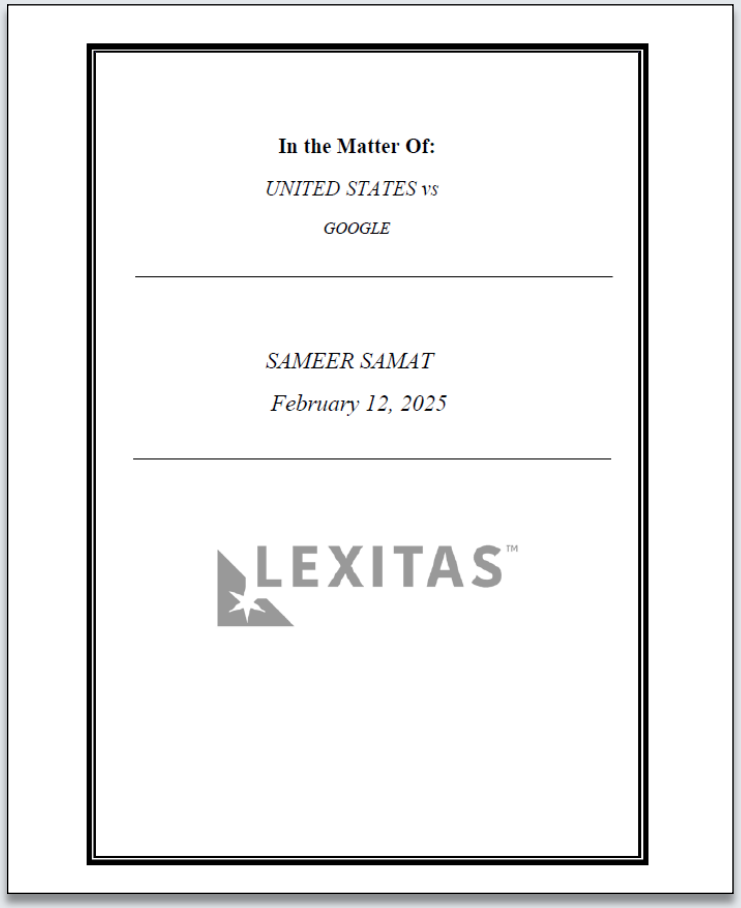
Google Is Aware of the Advantages AICore Can Give AI Models

Google unambiguously states in public-facing technical documentation that running models on AI accelerators via AICore is faster than running models outside of AICore on a phone's CPUs and GPUs.

"AICore is the new system-level capability introduced in Android 14 to provide Gemini-powered solutions for high-end devices, including integrations with the latest ML accelerators, use-case optimized LoRA adapters, and safety filters."

Source: PXR0332

Android's Lead Executive Confirms AICore's Exclusivity



In the Matter Of:

UNITED STATES vs

GOOGLE

SAMEER SAMAT

February 12, 2025

LEXITAS™

Q. But sitting here today, are you aware of any ways for a developer to access a device resident version of Gemini Nano other than through AICore?

A. I'm not. But I'm also not following all the ways that they are -- they may be doing that, and there could -- they could exist.

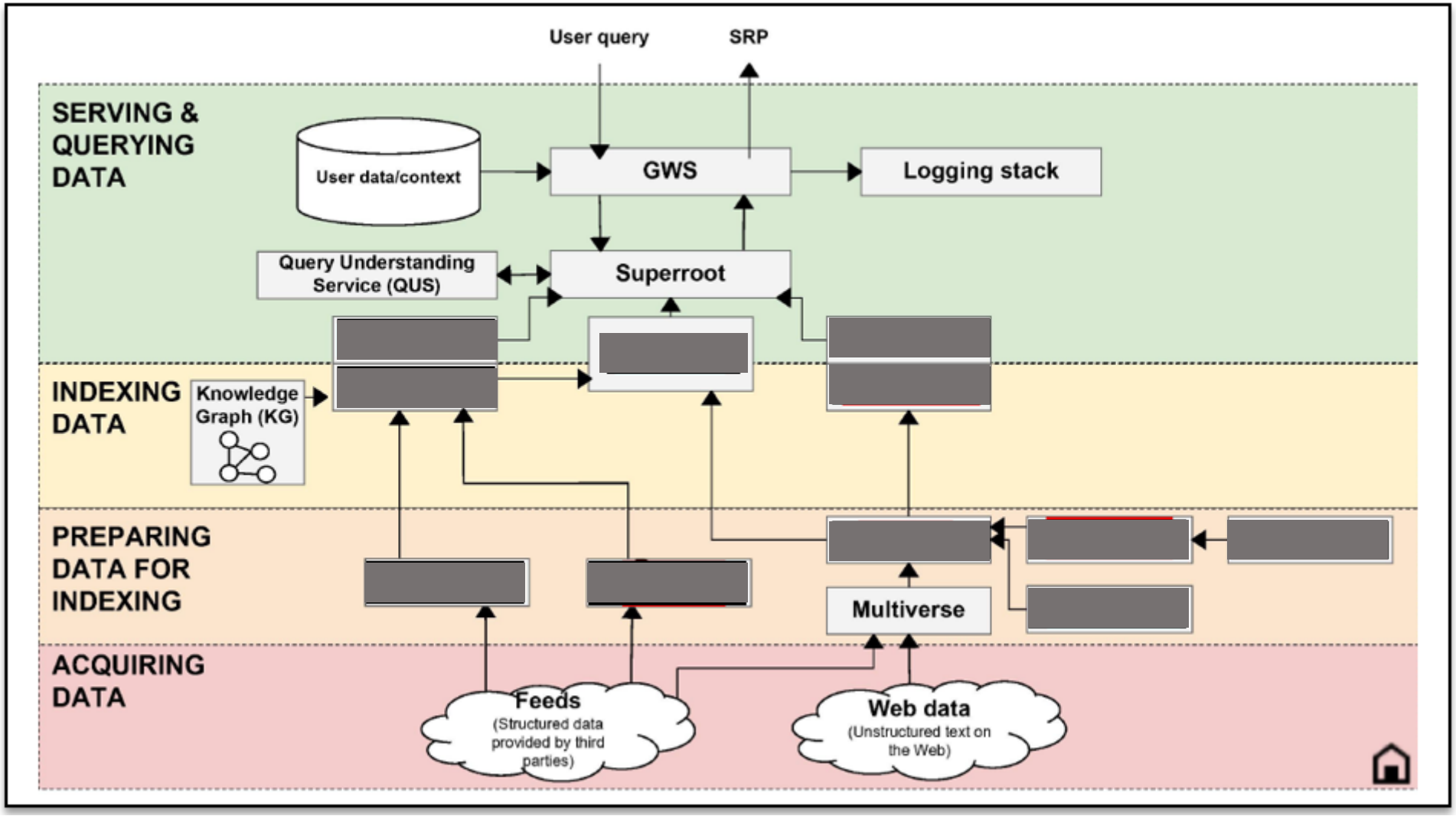
* * *

Q. Can AICore host other device-resident LLM models?

A. Today it cannot. We simply haven't gotten to that part of our road map yet.

Source: Samat Dep. 269:18-23; 270:3-6

Combined Search Infrastructure



Source: PXR0315 at -867

Unregulated AI Models Pose Security Risk

GUILLOTINE: Hypervisors for Isolating Malicious AIs

James Mickens
mickens@g.harvard.edu
Harvard University

Sarah Radway
sradway@g.harvard.edu
Harvard University

Ravi Netravali
rnetravali@cs.princeton.edu
Princeton University

ACM Reference Format:
James Mickens, Sarah Radway, and Ravi Netravali. 2025. GUILLOTINE: Hypervisors for Isolating Malicious AIs. In *Workshop on Hot Topics in Operating Systems (HOTOS '25), May 14–16, 2025, Banff, AB, Canada.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3713082.3730391

Abstract

As AI models become more embedded in critical sectors like finance, healthcare, and the military, their inscrutable behavior poses ever-greater risks to society. To mitigate this risk, we propose Guillotine, a hypervisor architecture for sandboxing powerful AI models—models that, by accident or malice, can generate existential threats to humanity. Although Guillotine borrows some well-known virtualization techniques, Guillotine must also introduce fundamentally new isolation mechanisms to handle the unique threat model posed by existential-risk AIs. For example, a rogue AI may try to introspect upon hypervisor software or the underlying hardware substrate to enable later subversion of that control plane; thus, a Guillotine hypervisor requires careful co-design of the hypervisor software and the CPUs, RAM, NIC, and storage devices that support the hypervisor software, to thwart side channel leakage and more generally eliminate mechanisms for AI to exploit reflection-based vulnerabilities. Beyond such isolation at the software, network, and microarchitectural layers, a Guillotine hypervisor must also provide physical fail-safes more commonly associated with nuclear power plants, avionic platforms, and other types of mission-critical systems. Physical fail-safes, e.g., involving electromechanical disconnection of network cables, or the flooding of a datacenter which holds a rogue AI, provide defense in depth if software, network, and microarchitectural isolation is compromised and a rogue AI must be temporarily shut down or permanently destroyed.

HOTOS '25, May 14–16, 2025, Banff, AB, Canada

ACM ISBN 979-8-4007-1475-7/2025/05
https://doi.org/10.1145/3713082.3730391

1 MOTIVATION

A machine learning model tries to emulate human reasoning. To do so, a model encodes observations about training data using numerical parameters and links between those parameters. Current state-of-the-art models are so large that their internal organization is opaque to their human creators. For example, the open-source BLOOM model has 176 billion parameters [27], the open-source Llama 3.1 model has 405 billion parameters [38], and the closed-source GPT-4 model is rumored to have more than a trillion parameters [58].

Humans cannot directly understand the relationships between such vast constellations of parameters. Automated methods for understanding those relationships (and how they generate model outputs) are an active area of research. Unfortunately, such model interpretability techniques appear to be inherently fragile. Consider the task of explaining LLM inferences. The soundness of LLM interpretability techniques is vulnerable to instabilities in the underlying LLM itself [60]. For example, the fact that LLMs are sensitive to minor changes in prompt phrasing [56] can result in a model's self-reported chain-of-thought being an unfaithful representation of the model's actual reasoning process [68]. Furthermore, an LLM's tendency to hallucinate [64] can manifest itself not only in the model's answer to a question, but in the model's explanation for that answer [80].

The opacity of model reasoning is troubling because models are increasingly connected to societally important infrastructure. For example, in financial settings, misbehaving models can generate huge monetary losses due to bugs. Those bugs might have been unintentionally introduced by model makers [41] or intentionally induced by adversarial examples [78]. In warfighting scenarios, military leaders are already concerned that AI-governed weapons may escalate conflicts due to ignorance of geopolitical nuance [42]; these escalation problems are exacerbated when AI makes decisions too quickly for humans to review those decisions [33].

Model alignment techniques [77] try to ensure that models adhere to human-defined behavioral norms. However (and concerningly), models can fake alignment compliance during training to later act in non-aligned ways post-deployment [21]. Researchers have also demonstrated that, in the specific context of LLMs, if model alignment does not completely eliminate the possibility of an undesirable model behavior, an adversarial prompt can always elicit that behavior [77]. Thus, society faces an increasing risk that an artifical general intelligence (AGI) model which matches or exceeds human reasoning will generate catastrophic harms in real life.

Abstract

As AI models become more embedded in critical sectors like finance, healthcare, and the military, their inscrutable behavior poses ever-greater risks to society. To mitigate this risk, we propose Guillotine, a hypervisor architecture for sandboxing powerful AI models—models that, by accident or malice, can generate existential threats to humanity. AI-

Source: James Mickens, Sarah Radway, and Ravi Netravali. 2025. Guillotine: Hypervisors for Isolating Malicious AIs. In Workshop on Hot Topics in Operating Systems (HOTOS '25), May 14–16, 2025, Banff, AB, Canada. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3713082.373039

ChromeOS and Aluminium Architecture

Tech Stack

Aluminium ('ChromeOS' built on the Android stack) will leverage Android's pre-existing stack.

Initial high-level gaps and dependency analysis.

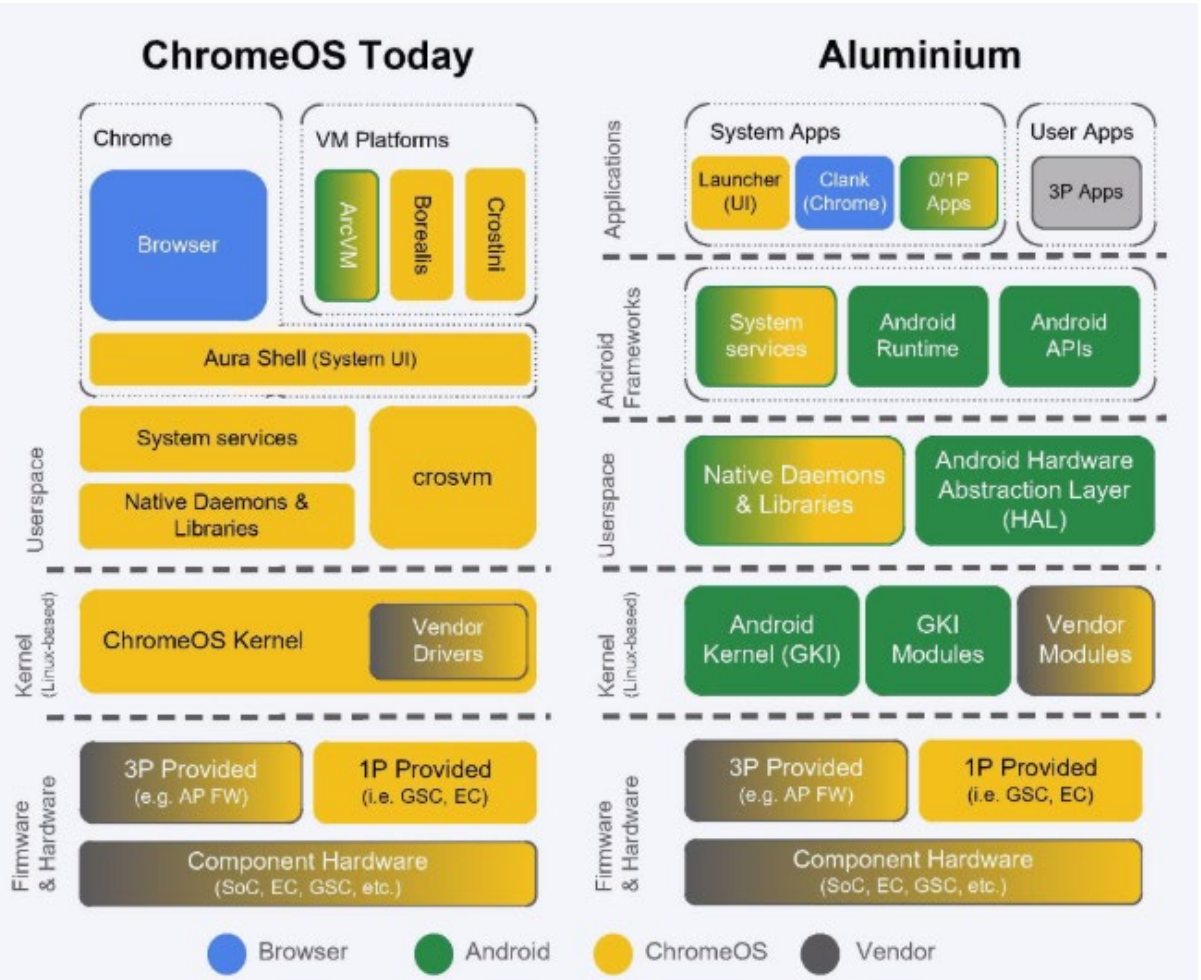
Through targeted ownership of key components AL would retain:

10 years of Google-provided updates

Security integrity from the hardware through the OS

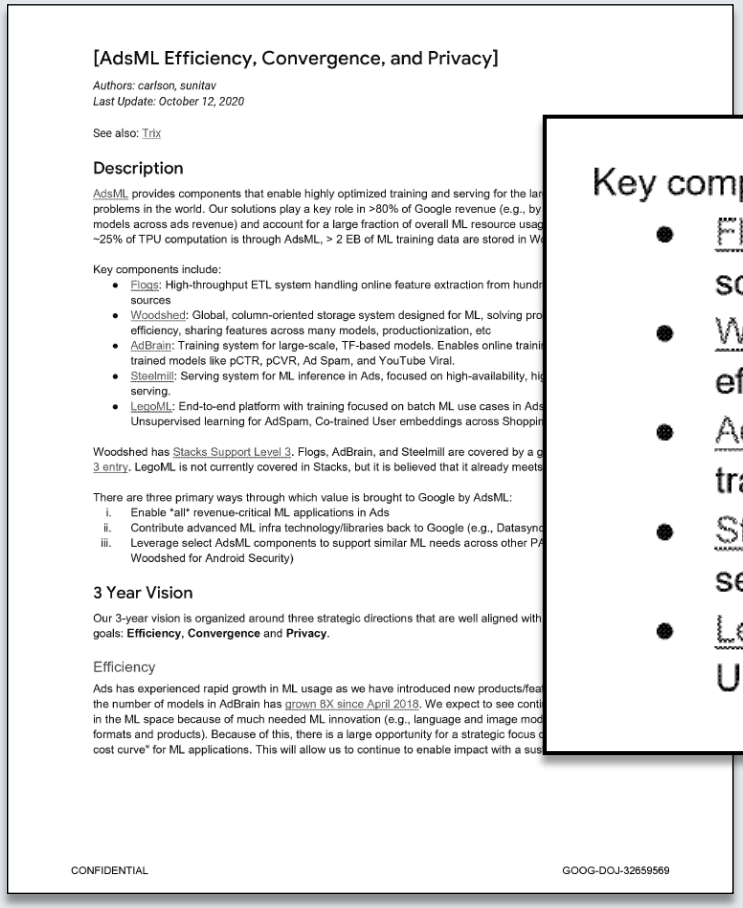
Desktop look and feel

ChromeOS mgmt layer



Source: PXR0317 at -420

Search Ads Machine Learning Infrastructure



[AdsML Efficiency, Convergence, and Privacy]

Authors: carlson, sunitav
Last Update: October 12, 2020

See also: Trix

Description

AdsML provides components that enable highly optimized training and serving for the lar
problems in the world. Our solutions play a key role in >80% of Google revenue (e.g., by
models across ads revenue) and account for a large fraction of overall ML resource usag
~25% of TPU computation is through AdsML, > 2 EB of ML training data are stored in W

Key components include:

- Flogs: High-throughput ETL system handling online feature extraction from hundr
  sources
- Woodshed: Global, column-oriented storage system designed for ML, solving pro
  efficiency, sharing features across many models, productionization, etc
- AdBrain: Training system for large-scale, TF-based models. Enables online traini
  trained models like pCTR, pCVR, Ad Spam, and YouTube Viral.
- Steelmill: Serving system for ML inference in Ads, focused on high-availability, hig
  serving.
- LegoML: End-to-end platform with training focused on batch ML use cases in Ads
  Unsupervised learning for AdSpam, Co-trained User embeddings across Shoppin

Woodshed has Stacks Support Level 3. Flogs, AdBrain, and Steelmill are covered by a g
3 entry. LegoML is not currently covered in Stacks, but it is believed that it already meets

There are three primary ways through which value is brought to Google by AdsML:

i. Enable *all* revenue-critical ML applications in Ads
ii. Contribute advanced ML infra technology/libraries back to Google (e.g., Datasync
iii. Leverage select AdsML components to support similar ML needs across other PA
   Woodshed for Android Security)

3 Year Vision

Our 3-year vision is organized around three strategic directions that are well aligned with
goals: **Efficiency**, **Convergence** and **Privacy**.

Efficiency

Ads has experienced rapid growth in ML usage as we have introduced new products/fea
the number of models in AdBrain has grown 8X since April 2018. We expect to see conti
in the ML space because of much needed ML innovation (e.g., language and image mod
formats and products). Because of this, there is a large opportunity for a strategic focus
cost curve" for ML applications. This will allow us to continue to enable impact with a sus

CONFIDENTIAL GOOG-DOJ-32659569

Key components include:

- Flogs: High-throughput ETL system handling online feature extraction from hundreds of raw data sources
- Woodshed: Global, column-oriented storage system designed for ML, solving problems around efficiency, sharing features across many models, productionization, etc
- AdBrain: Training system for large-scale, TF-based models. Enables online training for continuously trained models like pCTR, pCVR, Ad Spam, and YouTube Viral.
- Steelmill: Serving system for ML inference in Ads, focused on high-availability, high-performance serving.
- LegoML: End-to-end platform with training focused on batch ML use cases in Ads. Examples: Unsupervised learning for AdSpam, Co-trained User embeddings across Shopping & Search models.

Source: PXR0316 at -569