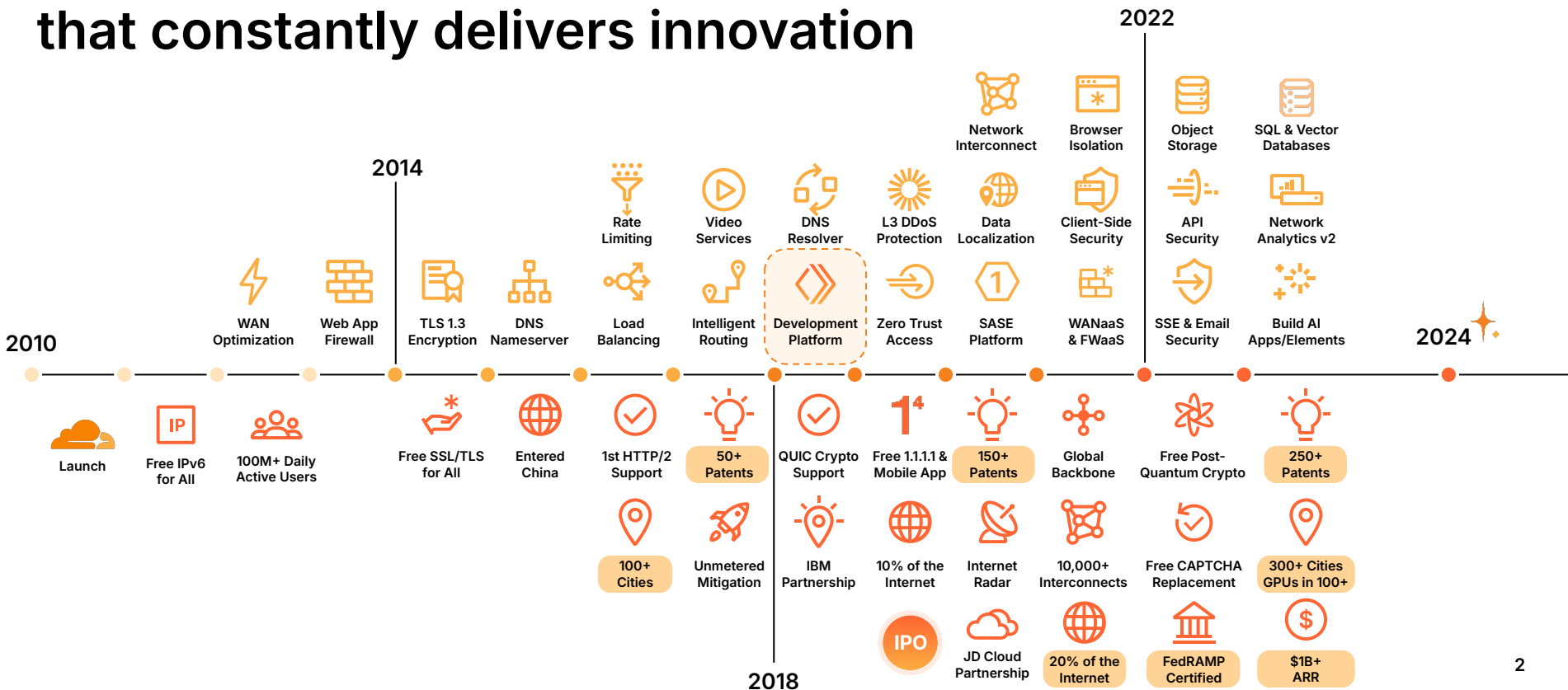# Cloudflare Developer Platform Overview

Lambros Petrou
Principal Systems Engineer, Developer Platform

Twitter @lambrospetrou / Bluesky @lambrospetrou.com
github.com/lambrospetrou

Agenda

1. Platform overview

2. Core components
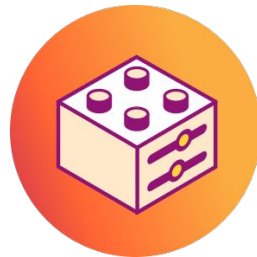
3. Gradual adoption

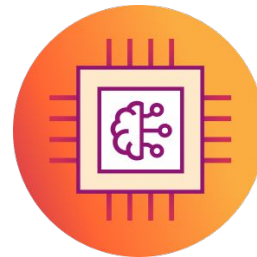4. Real-world examples and demos

Goals for Developers

Remove friction & improve developer velocity
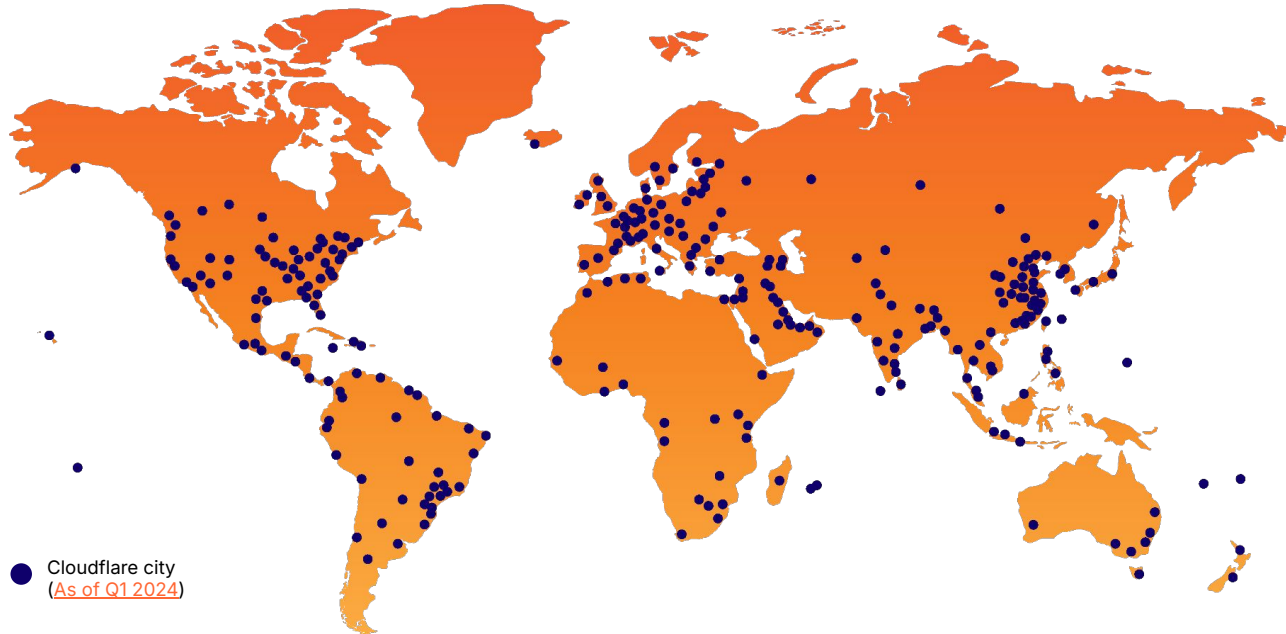
Provide primitives to build on

Enterprise grade solutions and scale

Enable new AI workloads

CLOUDFLARE

Cloudflare is the only composable, Internet-native platform...

...that delivers local capabilities with global scale.

Cloudflare city
(As of Q1 2024)

320
cities in 120+ countries,
including mainland China

150+
AI inference locations powered by GPUs

13,000+
networks directly connect to Cloudflare,
including every major ISP, cloud
provider, and enterprise

280 Tbps
global network edge capacity,
consisting of transit connections,
peering and private network
interconnects

~50 ms
from 95% of the world's
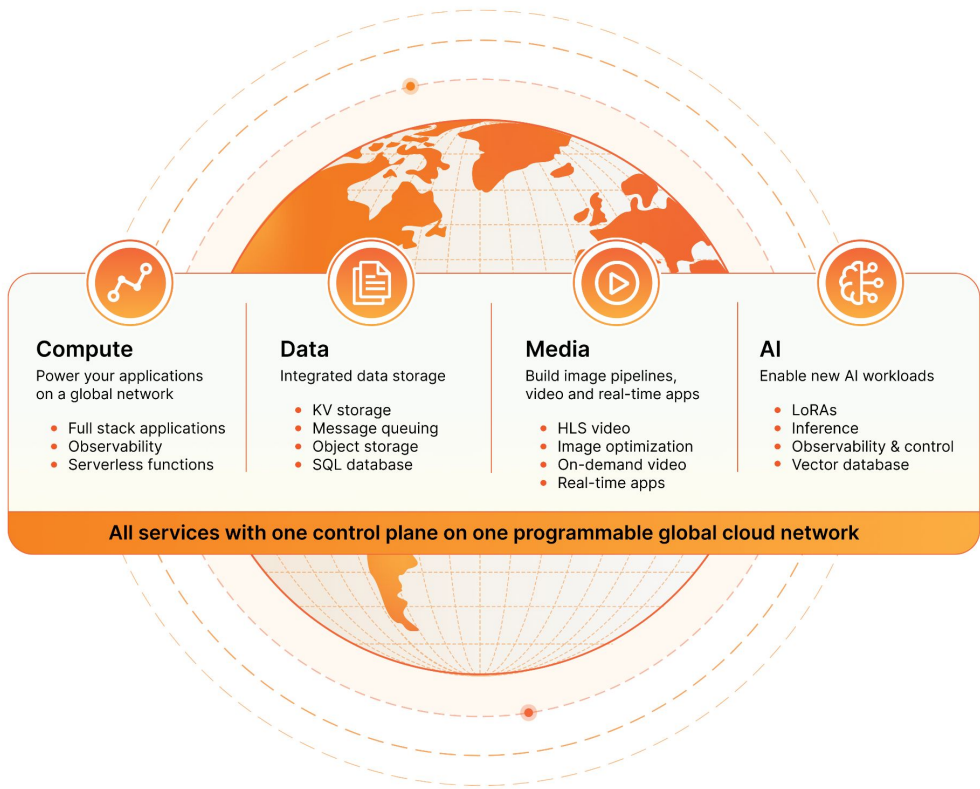Internet-connected population

5

Deploy applications to region: Earth*

One unified platform for building full-stack and AI applications uniquely powered by an intelligent, programmable global cloud network.

This enables organizations to ...

✓ **regain control,**

✓ **lower costs, and**

✓ **reduce the risks**

... of a complex and disjointed IT environment.

Compute
Power your applications on a global network

- Full stack applications
- Observability
- Serverless functions

Data
Integrated data storage

- KV storage
- Message queuing
- Object storage
- SQL database

Media
Build image pipelines, video and real-time apps

- HLS video
- Image optimization
- On-demand video
- Real-time apps

AI
Enable new AI workloads

- LoRAs
- Inference
- Observability & control
- Vector database

All services with one control plane on one programmable global cloud network

Supporting developers through the full SDLC

Local Development

Testing

Deploy

Observe

CLOUDFLARE

Developers who know, trust Cloudflare

Delivery Hero

100+
of hours saved in development and testing time with Cloudflare Workers.

Caliente

$1000+
in monthly savings when switching to R2 for object storage.

PORSCHE INFORMATIK

40%
improvement in performance with a globally distributed asset cache and Workers powering the application

Dataweavers

" Workers and Cloudflare Image Optimization allow us to do much of our processing right at the edge, without roundtrips to origin servers. That reduces costs and improves the end-user experience by speeding up frequent requests, authentications, and image processing and delivery tasks."

- Ben Shapiro, CEO, Dataweavers

Built with stories

CLOUDFLARE

What makes us different

Developer experience

Performance built-in

Infinitely scalable

No trade offs
No need to choose between cost, performance and security.

Limitless interoperability
Adopt new capabilities in any order Integrate with any 3rd party tool

Save time
No tuning necessary. Automatic placement for optimal performance.

Zero cold starts
Isolate vs container architecture eliminates cold starts.

Deploy from region: Earth
Code executes within 50ms of ~95% of the Internet-connected global population.
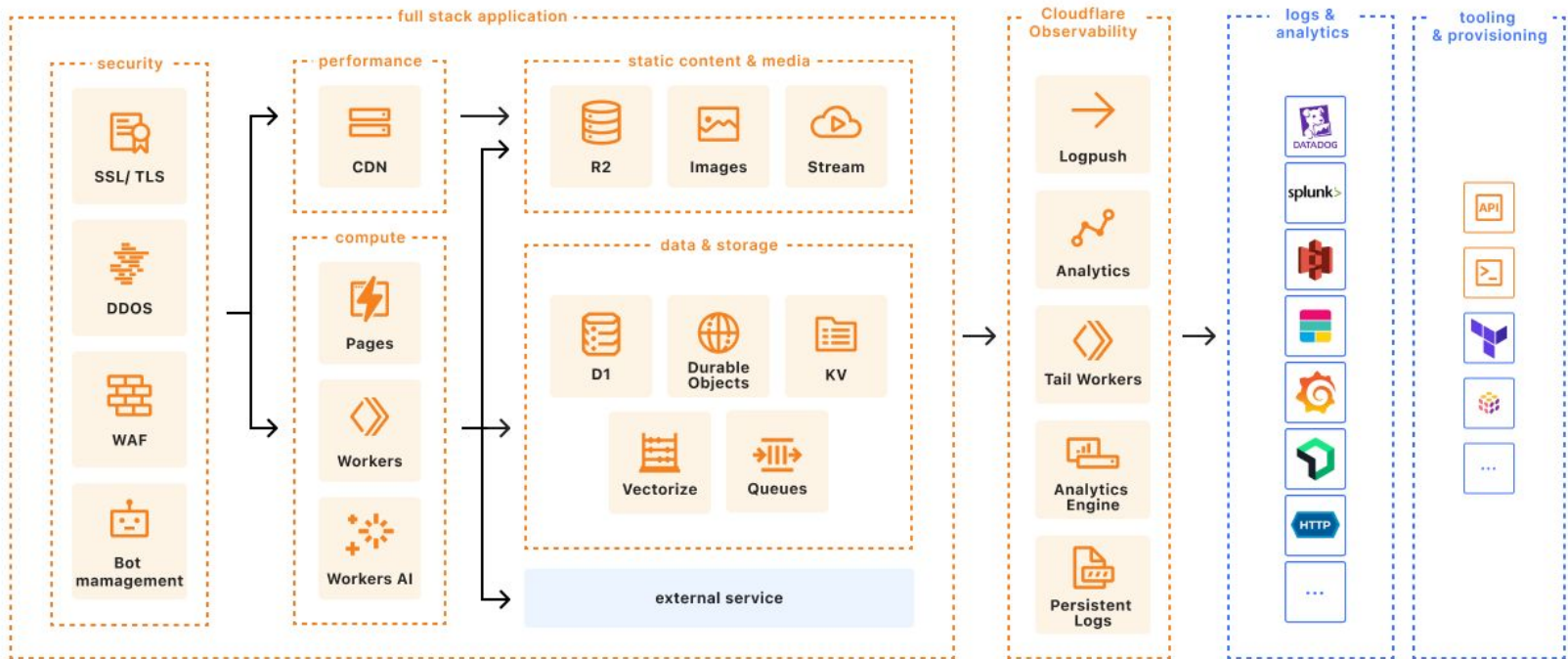
Flexible pricing

With zero egress fee object storage and compute pricing based on CPU time no need to worry about excessive costs as you scale

Join the 2 million developers building on Cloudflare

Core components of the platform

Full-stack application reference architecture

New Reference Architecture center

Cloudflare Workers

- Serverless compute, based on V8 isolates running JavaScript code or WASM. Python support in beta.

- Integrated within the entire Cloudflare network, with dedicated runtime APIs making most of Cloudflare developer platform products available inside your Worker code, like direct access to the CDN cache.

- Fast cold starts, running everywhere on Cloudflare's network.

- References

  - https://developers.cloudflare.com/workers/reference/how-workers-works/

  - https://developers.cloudflare.com/workers/reference/how-the-cache-works/

Cloudflare Durable Objects

- Durable Objects combine compute, durable storage across requests, and Cloudflare's edge network. Actors at the infrastructure level.

  - Read Zero-latency SQLite storage in every Durable Object blog post.

- Each Durable Object (instance) is a mini server with 128MB of memory, running your Worker code, with persisted durable storage (10GB) with KV or SQLite API.

- Create millions of them, programmatically! 🚀

- Each Durable Object can be created in the "region" of your choice or by default close to your request location.

  - https://developers.cloudflare.com/durable-objects/reference/data-location/

  - https://where.durableobjects.live/

13

Cloudflare Durable Objects

- https://github.com/lambrospetrou/tiddlyflare/blob/2f6cd98eab2d77f8319cca21922dea3a8ca41d9a/src/durable-objects.ts#L543

```
async function routeCreateWiki(env, tenantId, name, wikiType) {
    let id: DurableObjectId = env.TENANT.idFromName(tenantId);
    let tenantStub = env.TENANT.get(id, { locationHint: "weur" });
    const resp = await tenantStub.createWiki(tenantId, name, wikiType); ★★★
    return { redirectUrl: resp.redirectUrl };
}
```

Storage for any use case

Workers KV   `{ }`   Distributed, eventually consistent key-value store to serve read-heavy non-dynamic data like configurations and product information.

R2   S3 compatible, egress free, object storage to store images, media, logs, LLMs, and more.

Hyperdrive   Make your existing database feel globally distributed and connect directly from Workers.
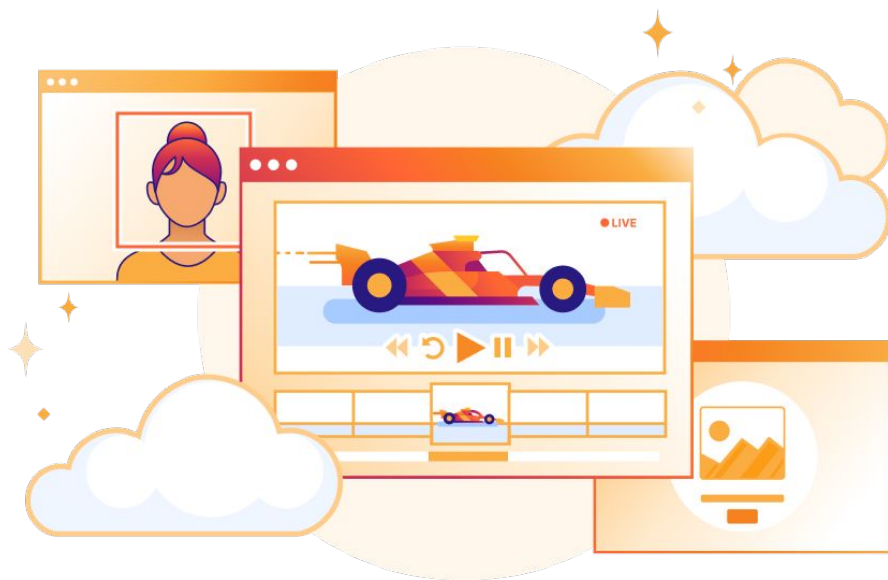
D1   Build serverless relational SQL databases for user, product, or other types of data with point in time recovery.

Durable Objects   Durable Objects combine compute, durable storage, and Cloudflare's edge network. Improve application performance by colocating both code and data close to users.

Create immersive multimedia experiences

- Deliver live and on-demand video

- Create an all-in-one image pipeline

- Build real-time video & audio applications

Elements of AI applications and infrastructure

Train

Store training data

Create and store model

Fine tune models

Secure

Secure access to AI services

Protect private data

Deploy

Generate embeddings

Deploy model
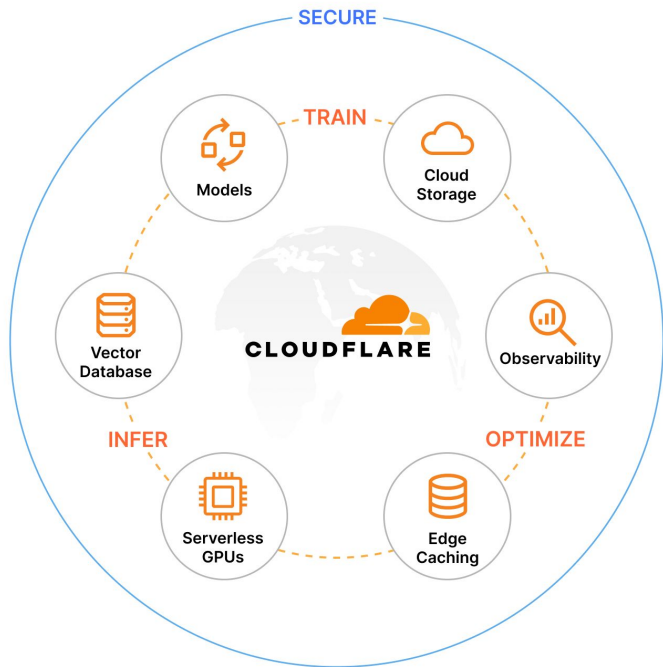
Run inference at scale

Optimize

Observe and analyze usage

Control traffic & optimize cost

A holistic, end-to-end AI solution

1) TRAIN
Zero Egress Cloud Storage with **R2**

2) INFER
Generate, store, and search embeddings with
Workers AI and **Vectorize**

3) OPTIMIZE
Gain visibility and improve performance
via **AI Gateway**

4) SECURE
Protect and Defend with **Zero Trust** and **API
Security**

18

Gradual adoption of the developer platform

CLOUDFLARE

Workers and CDN Cache

- Proxy your API with Workers and access the CDN Cache directly from your code to cache each response (or even partially) accordingly.

  - https://developers.cloudflare.com/workers/runtime-apis/cache/

  - https://developers.cloudflare.com/workers/reference/how-the-cache-works/

  - https://developers.cloudflare.com/workers/examples/cache-api/

- Smart placement

  - Place your Worker close to your origin(s) and route within Cloudflare.

  - https://developers.cloudflare.com/workers/configuration/smart-placement/

R2

- Eliminate egress fees by using R2.

- S3-compatible API for zero code changes in your existing applications.

- Worker-API for direct integration with your Worker code.

- https://r2-calculator.cloudflare.com/

Workers KV

- Improve performance using Workers KV for your read-heavy workloads.

- Eventually consistent globally within 60s.

- KV is a global, low-latency, key-value data store. It stores data in a small number of centralized data centers, then caches that data in Cloudflare's data centers after access.

- Example: `await env.KVNamespace01.put("first-key", "This is the value for the key");`

- References

  - https://developers.cloudflare.com/kv/concepts/how-kv-works/

Hyperdrive

- Use Hyperdrive to connect your traditional databases on other clouds or on-premises, while taking advantage of Workers.

- https://hyperdrive-demo.pages.dev/ (open this 😉)

- References

  - https://developers.cloudflare.com/hyperdrive/configuration/how-hyperdrive-works/

  - https://developers.cloudflare.com/hyperdrive/reference/supported-databases/

CLOUDFLARE

Workers AI

- Use Workers AI to invoke, fine-tune, and work with many LLMs through a single completely managed platform.

- References

  - https://developers.cloudflare.com/workers-ai/models/

  - https://developers.cloudflare.com/workers-ai/get-started/rest-api/

  - https://developers.cloudflare.com/workers-ai/configuration/open-ai-compatibility/

  - https://developers.cloudflare.com/workers-ai/configuration/bindings/

CLOUDFLARE

Cloudflare works with your existing infrastructure

- Use as much or as little from the Cloudflare Developer Platform as you need.

- Move workloads gradually, one by one.

- Develop new internal tools and products on Cloudflare for faster development, shipping features faster, and without worrying about regions and infrastructure.

- Cloudflare works well in front of any cloud provider, VPS/server providers, even with on-premises infrastructure.

Real-world examples and demos

Useful resources

- Control and data plane architectural pattern for Durable Objects

- Cloudflare built on Cloudflare

  - https://blog.cloudflare.com/how-we-built-cloudflare-queues/

  - https://blog.cloudflare.com/80-percent-lower-cloud-cost-how-baselime-moved-from-aws-to-cloudflare/

- https://www.lambrospetrou.com/articles/tiddlyflare/

- https://github.com/lambrospetrou/rediflare

27

CLOUDFLARE

Ensuring your success

Community	Docs	Resource Hub	Help Center	Discord
Blog	theNET	Cloudflare Radar	CloudflareTV	YouTube

CLOUDFLARE

Questions?
Comments?
Feedback?

Lambros Petrou
Principal Systems Engineer, Developer Platform

Twitter @lambrospetrou / Bluesky @lambrospetrou.com
github.com/lambrospetrou