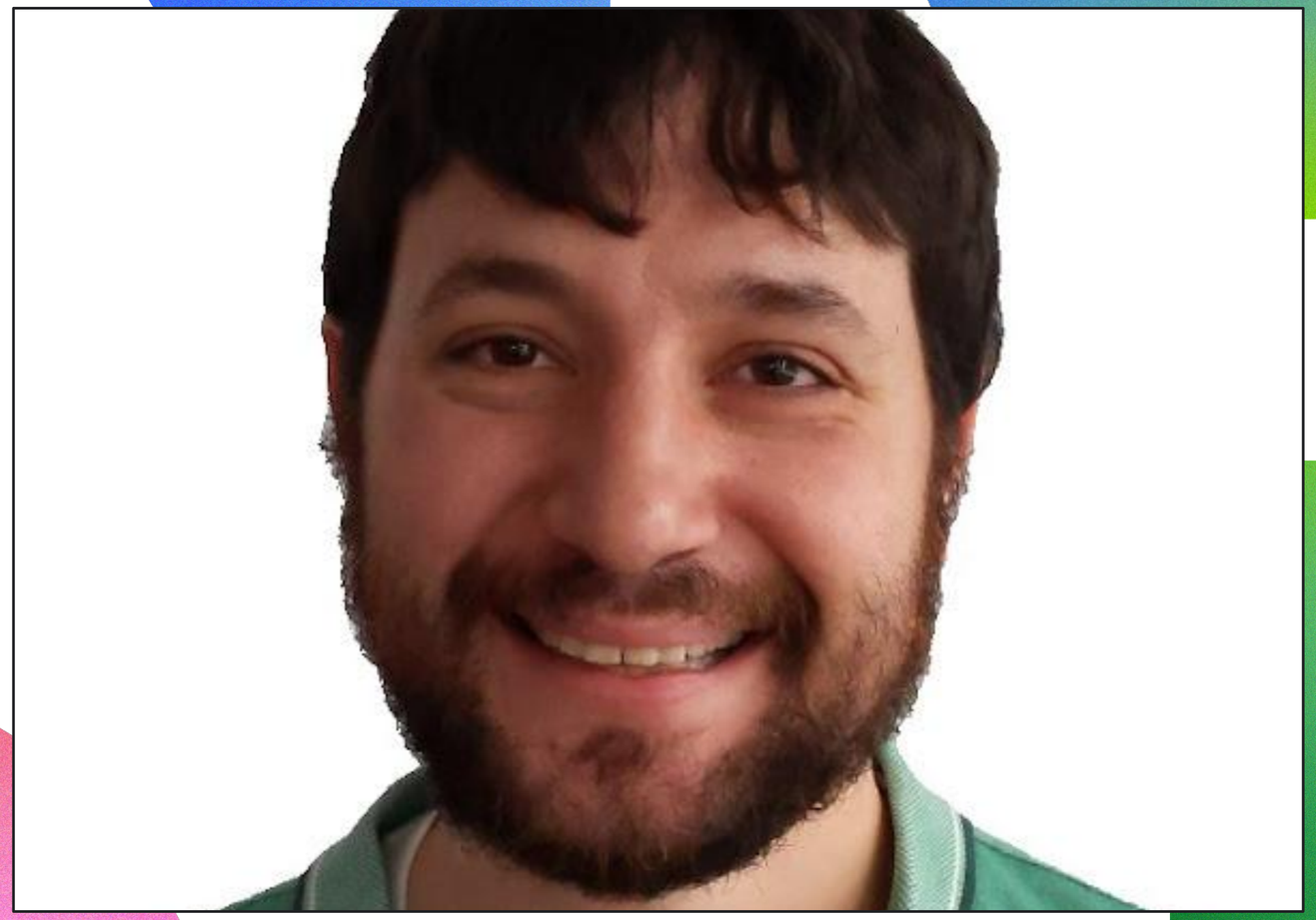
Google Cloud

Next '24

Cloud Run: what's New

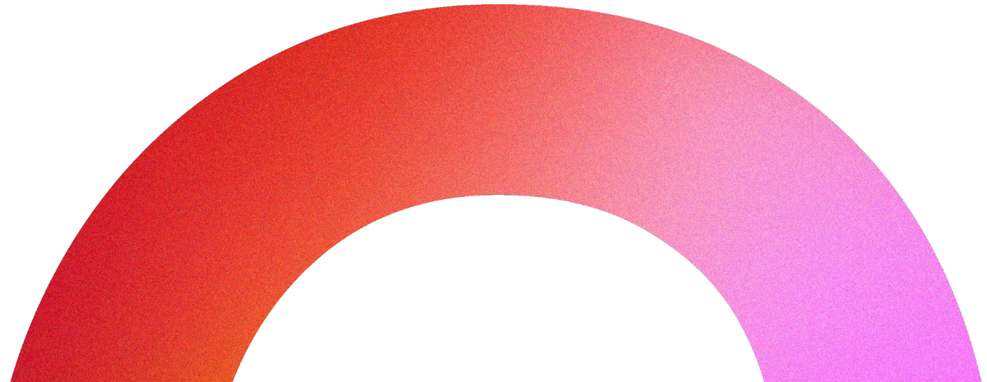
Justin Mahood

Product Manager, Google Cloud



Google Cloud's serverless runtime

Run applications fast and securely in a fully-managed environment.



Cloud Run

Experience

Simple
Demand as little as possible.

Automated
Cloud Run takes care of a lot for you.

Top satisfaction and usability scores
High satisfaction and task success.

Developer productivity
Idiomatic to developers, deployment velocity.

Runtime

Capable
Run any container.

On-demand
No pre-provisioning.

Hyper-elastic
Fast automatic scaling, including to zero.

No infrastructure management
No VM or cluster to manage.

Pricing

Pay only when code is running,
with a 100ms granularity.

- CPU
- Memory
- Requests (not always)

Perpetual monthly free tier

Committed Use Discounts

No flat fee!
"if you don't use it, you don't pay for it"

Agenda

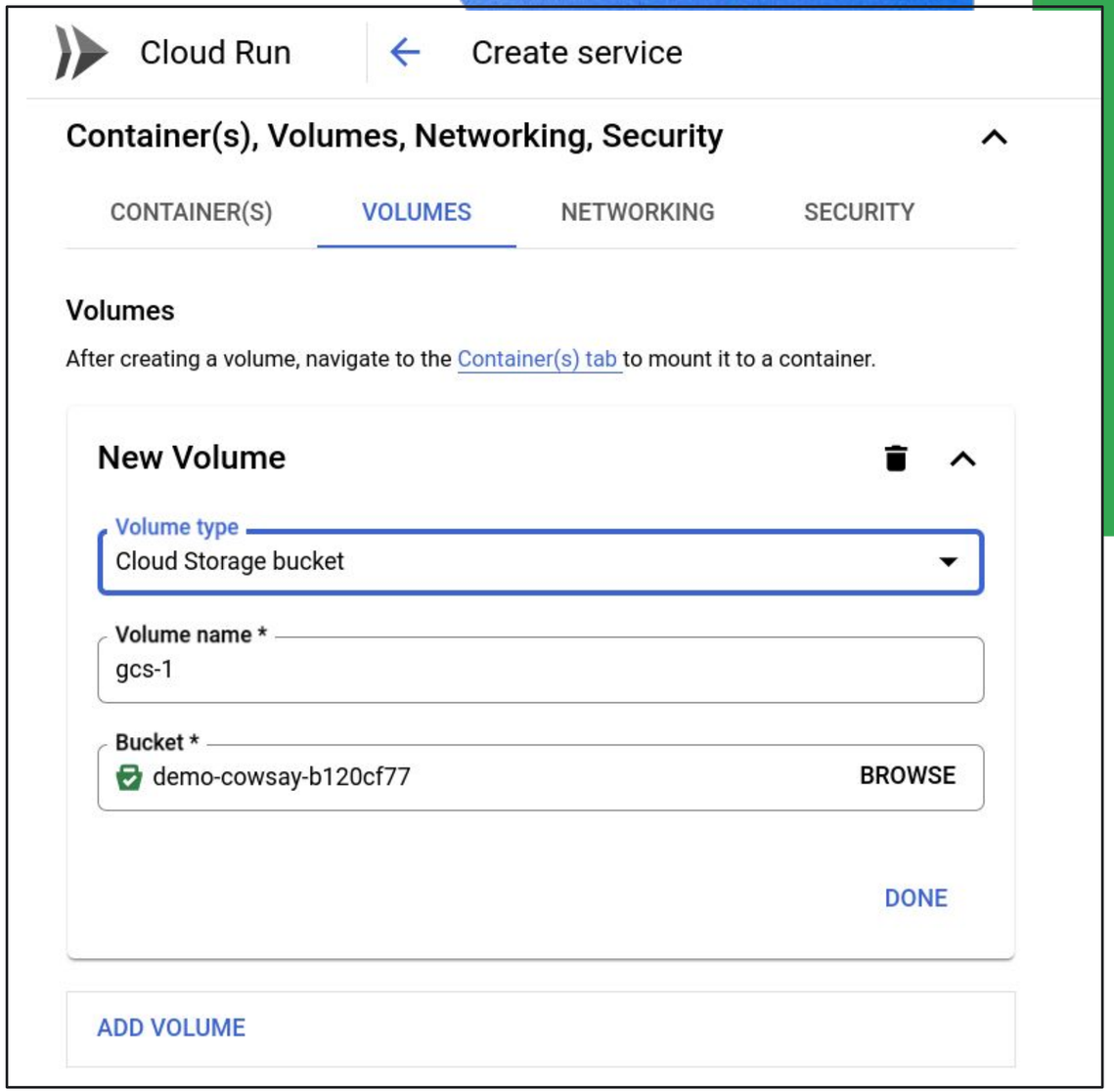
01 Simplifying App Development
02 Demos
03 Enterprise Ready
04 Customer Story: DZ BANK AG

Simplifying App Development

Volume Mounts

Preview

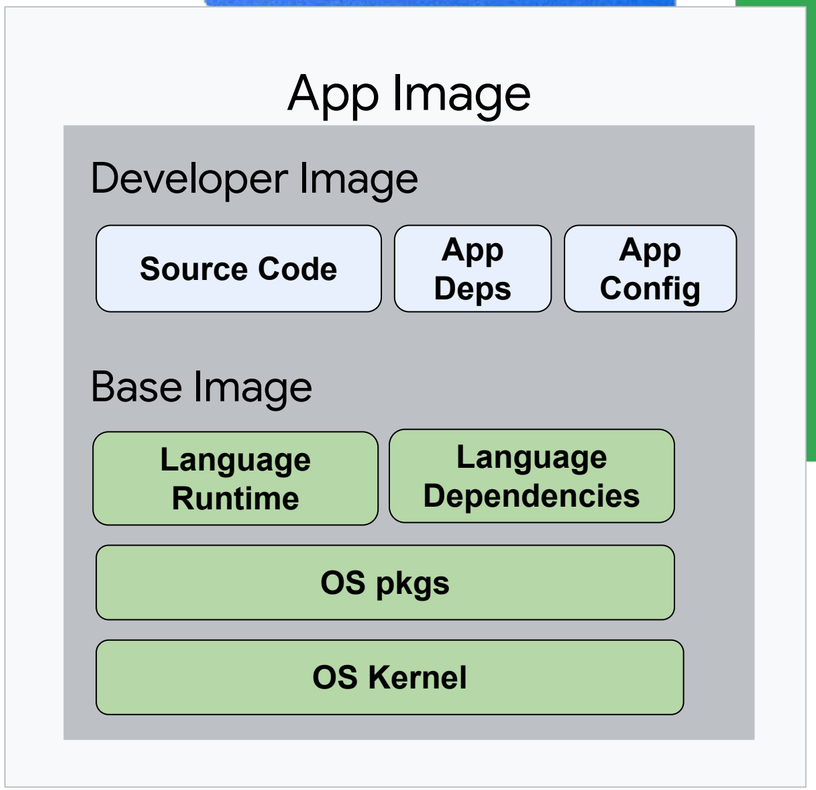
- Mount NFS or Cloud Storage Fuse filesystems as read-only or read-write
- Implemented through sidecar which handles kernel level file mount



Automatic Security Updates

Private Preview

- Automatic base image updates on deployed images
- 0 downtime, 0 rebuild
- Target < 48hr patch time to security incidents




```
> glcoud run deploy myapp --source .
 --base-image=us-central1-docker.pkg.dev/serverless/google-22/go:latest
```

Deterministic URL

Private Preview

- Access your Cloud Run service with a consistent URL scheme
- The leftmost DNS segment must be <= 63 characters long
- Coming soon: support for Private Google Access, and Private Service Connect
- Disable ‘run.app’ URL entirely if desired

https://myservice-1234567.us-central1.run.app

service name | project number | region

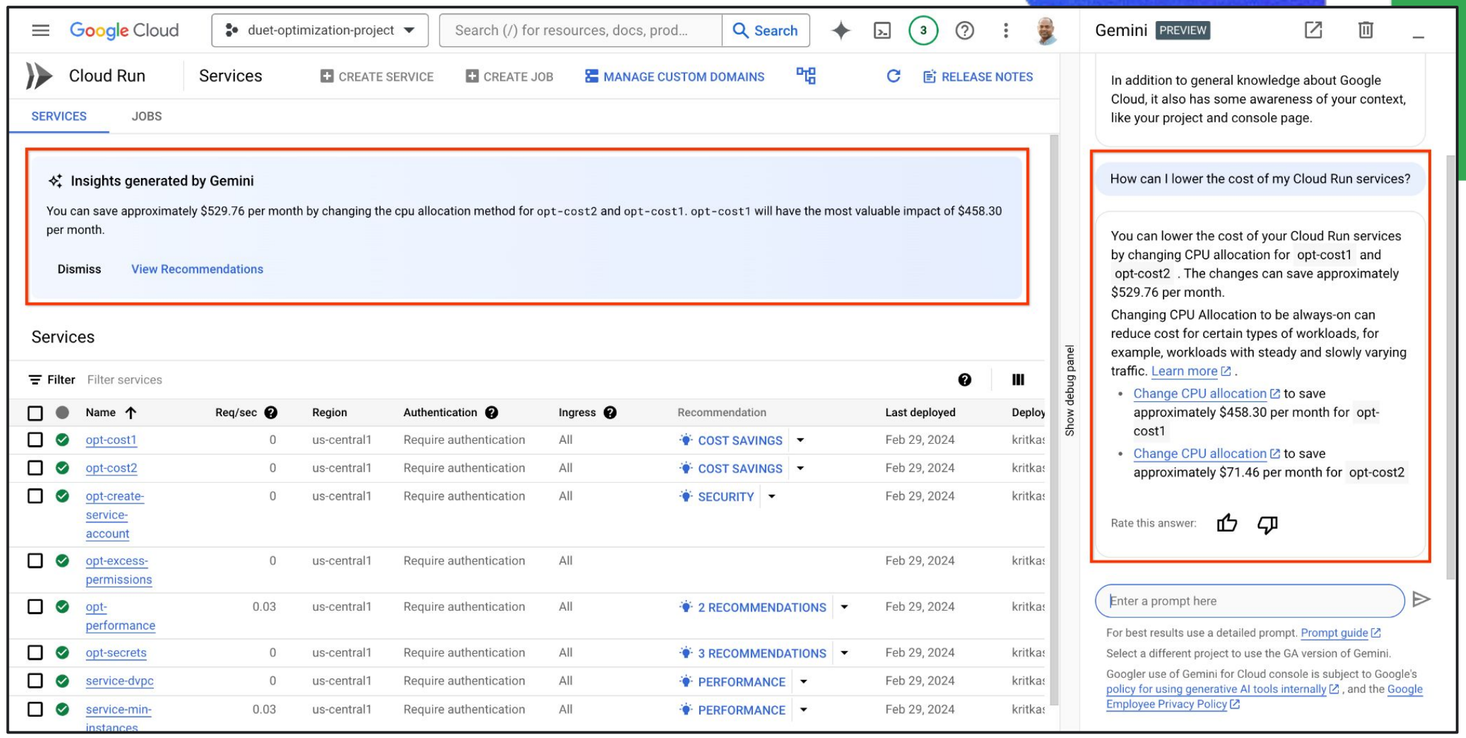
or

```
$ gcloud run deploy --no-default-url
```

Gemini in Cloud Run Recommendations

Private Preview

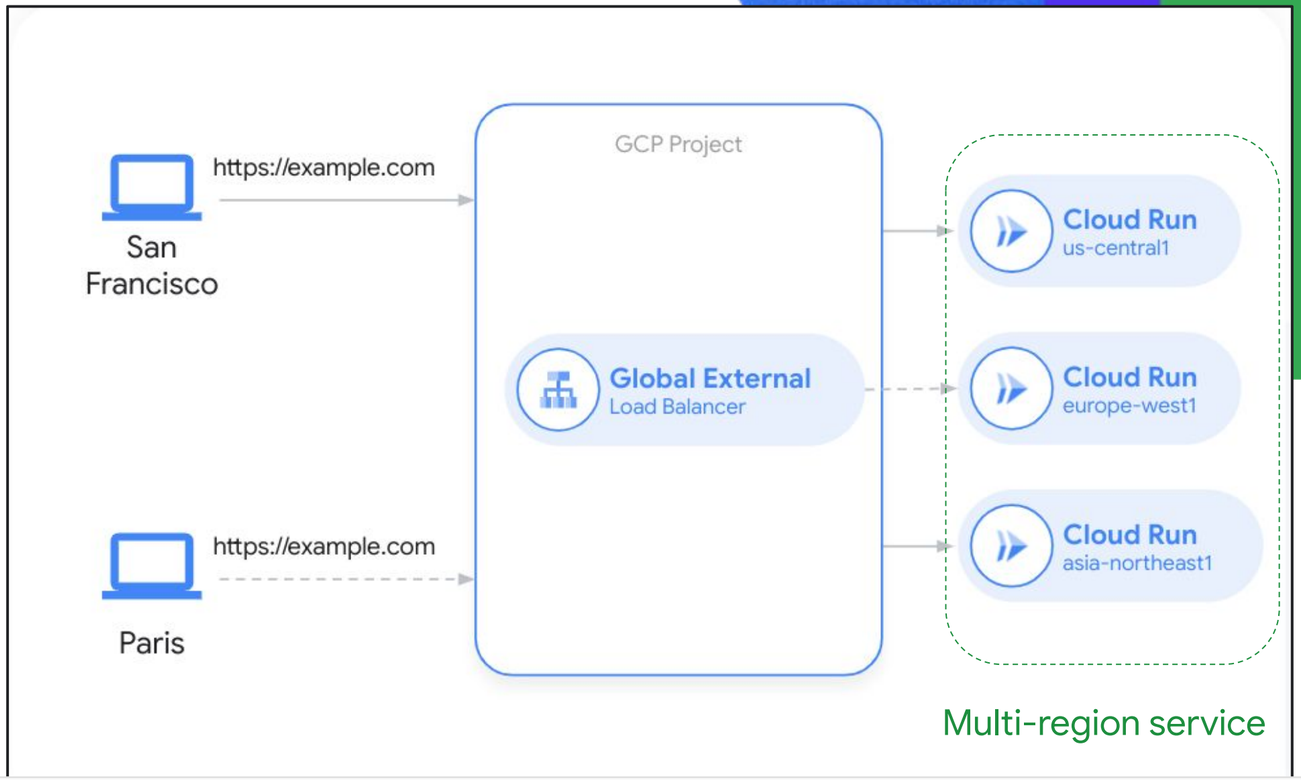
- Surface recommendations proactively with banners and reactively with chat
- Aggregate recommendations for the most impact
- Summarize recommendations for easier impact analysis



Multi-region services

Private Preview

- Deploy the same service to **multiple regions** with a **single command**
- Automatically expose a **global endpoint** that routes requests to the closest region




```
> glcoud run deploy myapp --region=us-central1,europe-west1,asia-east1
                          --domain=example.com
```

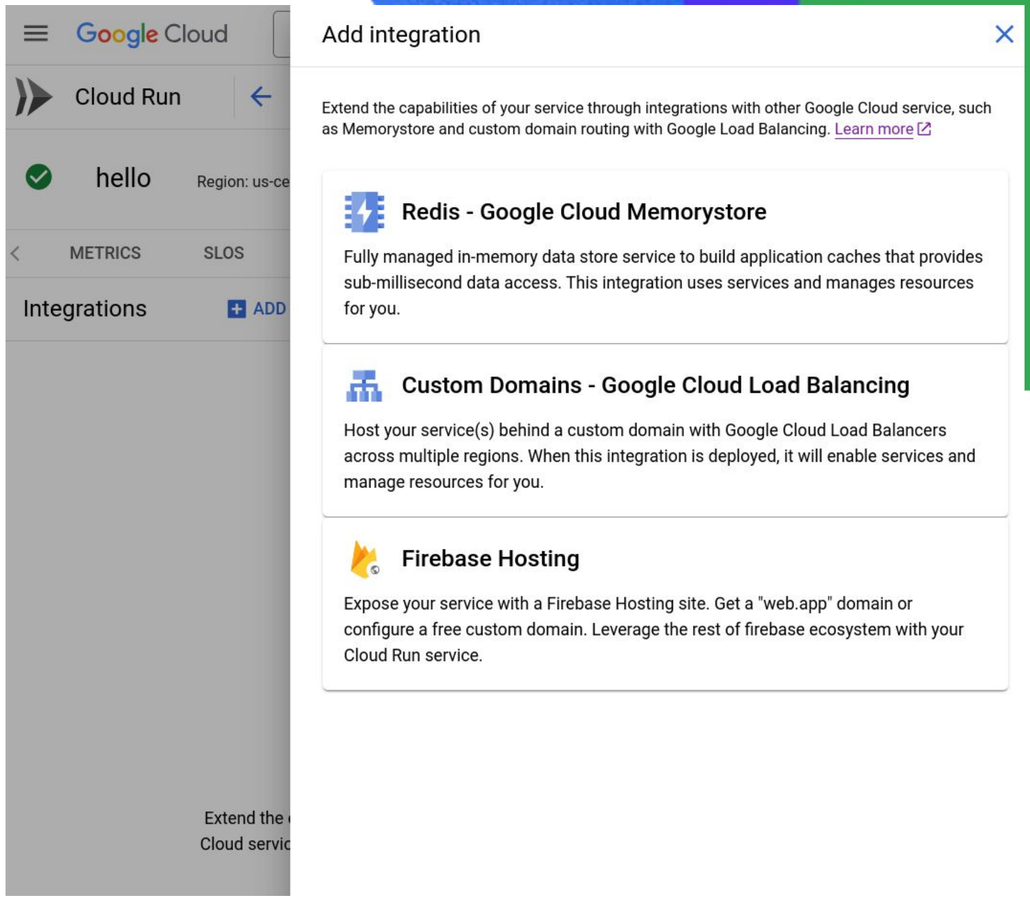
Demo

Simplifying Application Development

No application is just a container

Cloud Run Integrations

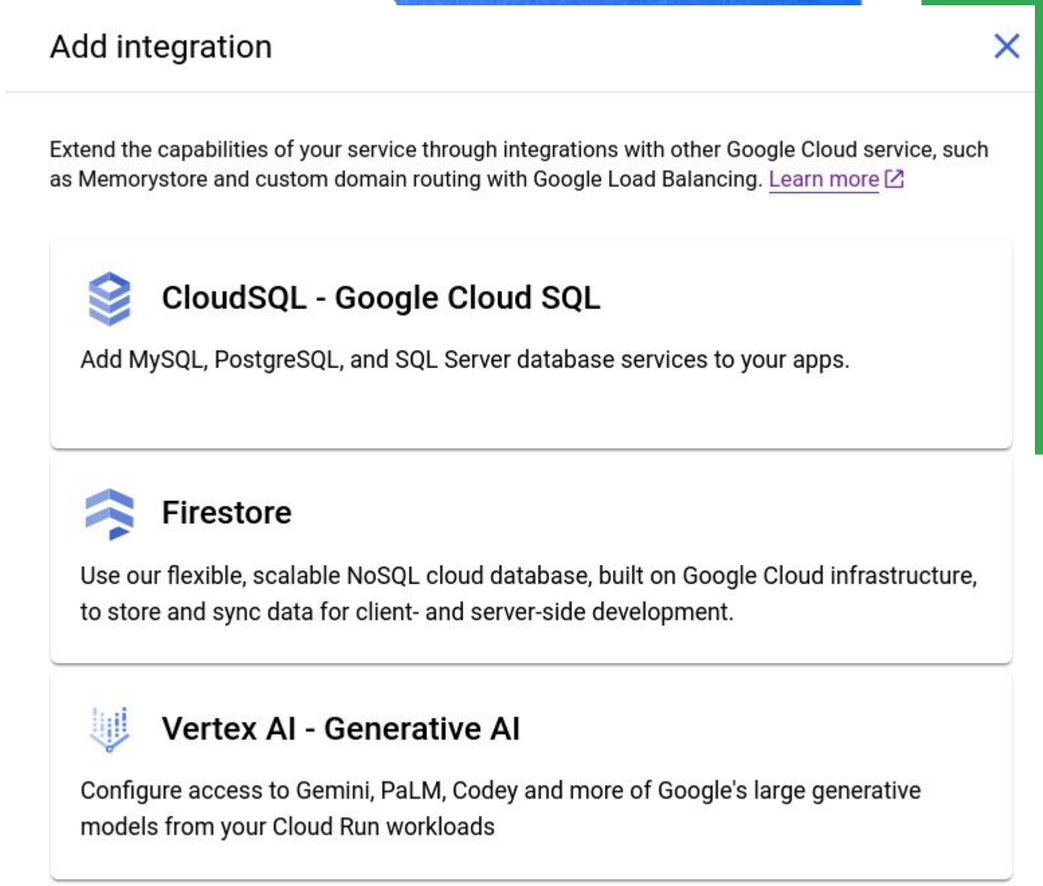
- One command for complex integrations
- Configures Cloud Run for you
- Easy access to all related resources



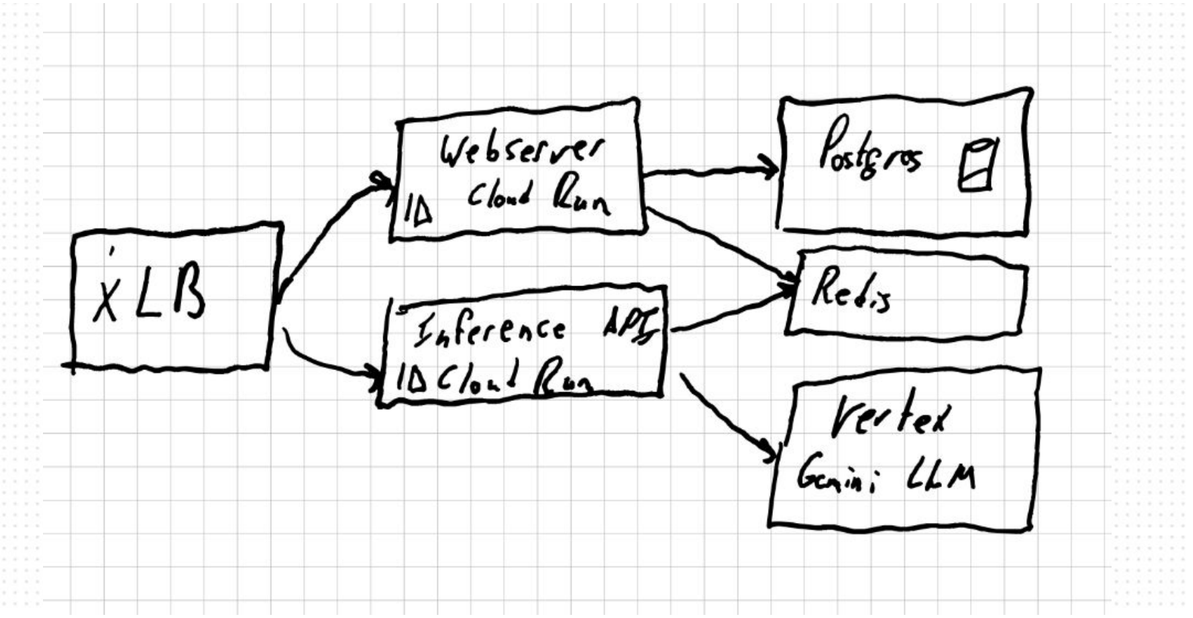
New integrations

Preview

- Three new Cloud Run integrations: CloudSQL, Firestore, and VertexAI
- ‘1 click’ to automatically configure and connect GCP services to your Cloud Run workloads



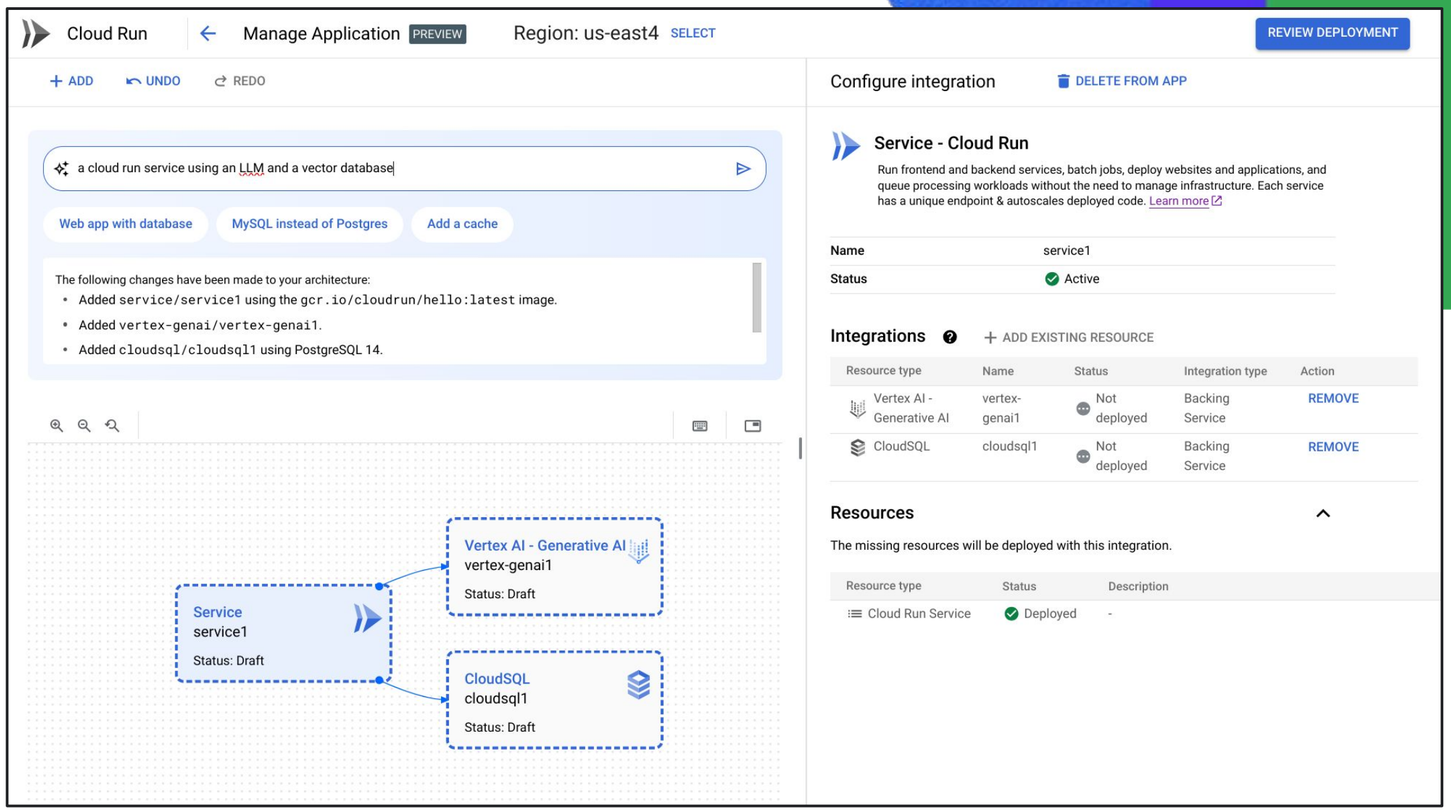
Integrations were just the first step....



Application canvas

Private Preview

- Simplify the creation of complex applications to a few clicks, and visualize in a diagram
- Use Gemini to edit or generate a **deployable** architecture using natural language



Demo

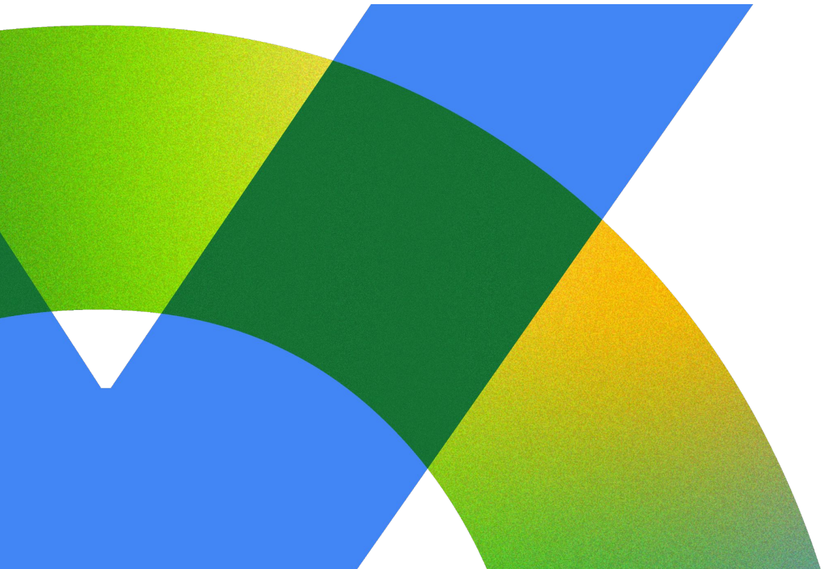
Enterprise Ready

Sridhar Venkatakrishnan

Engineering Manager,
Google Cloud

Enterprise workloads

Large enterprises have unique needs



- Advanced security and compliance needs
- Migrating from on-premises to cloud native containerized workloads
  - Slow starting containers (e.g. Java Spring Boot)
- Large scale
  - CPU and memory hungry
  - Many services and instances
  - Cost and performance sensitive
- Support complex network architectures spanning multiple domains

Secure

Access Control

IAM invoker permission

VPC Service Controls

Identity Aware Proxy

Supply Chain

Secure Software Supply Chain insights in Console

Automatic base image updates **(Private Preview)**

Binary Authorization

Encryption

Customer Managed Encryption Keys

Secret Manager integration

Compliance

ISO, SOC, PCI

FedRAMP Moderate, NIST, IL2

HIPAA

Assured Workload

Network Security

Ingress & egress controls

Firewall rules with Direct VPC Egress

Disable default run.app URL **(Private Preview)**

Cost effective at scale

Max Instances

Specify a maximum number of instances to limit the upper-bound of scaling

Always on CPU

"Always-on CPU" is priced 25% lower than "throttled CPU"

More economical for large scale workloads

Committed Usage Discounts

Self-service Committed Use Discounts offering a 17% reduction in usage cost.

Active Assist

Proactive recommendations to optimize pricing selection and commitment.

Announcing: Direct VPC Egress GA

Scalable

- 2x throughput compared to VPC Connectors
- Lower latency
- Lower cost

Compatible

- All Cloud Run regions supported*
- Cloud NAT*
- Firewall rules logging and VPC flow logs*

Simple

- Simple configuration
- Manage with org policy
- Instance limits are now a quota*

* New since Public Preview

Announcing:

Cloud Service Mesh with Cloud Run

Private Preview

Benefits

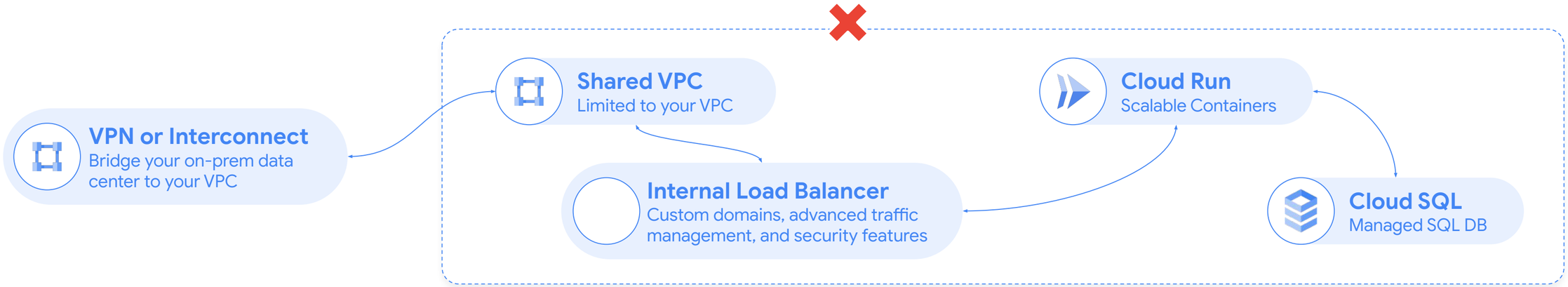
- Traffic management
- Observability
- Security

Notable Cloud Run features

- Fully-managed data plane (Envoy sidecars)
- Friendly service names
- JWT token injection

Integrated

Integrate easily with your existing networks and systems



(Optional) on-prem VM calling through **VPN** or **Interconnect**

Your **private shared VPC** may contain internal resources and users with a security boundary enforced at the network level

Internal Load Balancer gives you custom domains, advanced traffic management, and security features

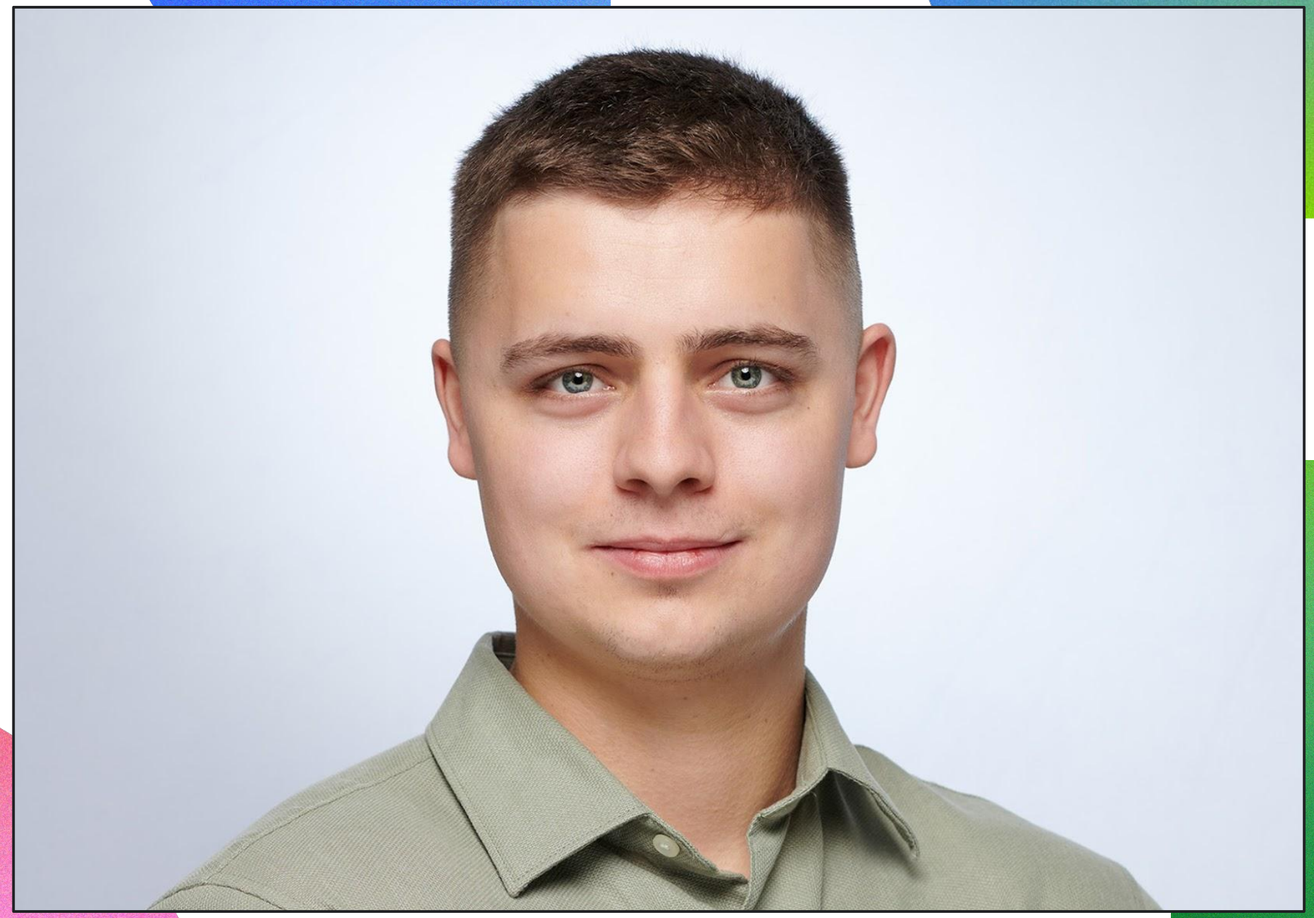
Cloud Run will only accept requests from within your project or shared VPC network, and will prevent egress to any destination outside the VPC

Other Google Cloud resources within the VPC boundary are accessible

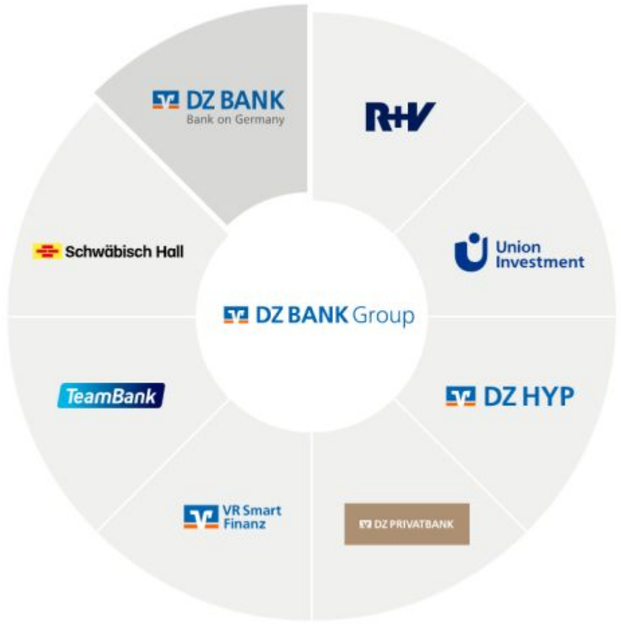
Customer Story: DZ BANK AG

Tim Harpe

Senior Cloud Engineer, DZ Bank



DZ BANK



Industry: Financial Services
Country: Germany

Customer Award

Harissa
Market price risk application build on GKE

ESG Data
ESG Application in our credit processes build on Cloud Run

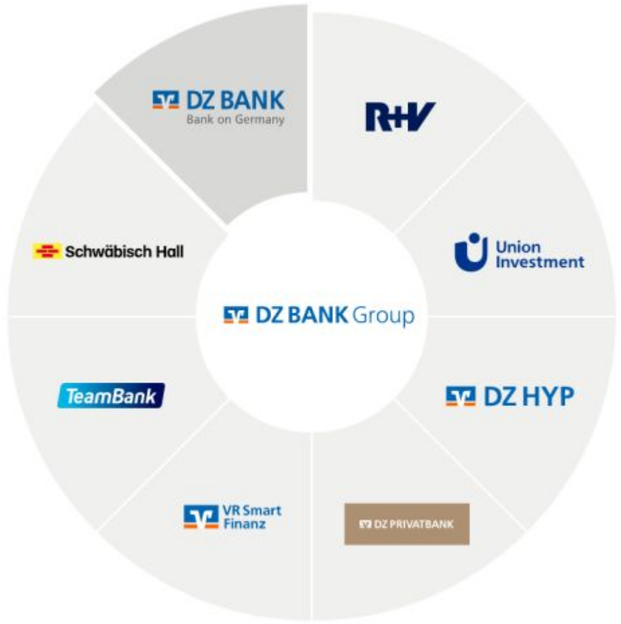
2020 — 2021 — 2022 — 2023 — 2024

Landing Zone
Productive use

Flake
Custom data lake powered by BigQuery

Developer Platform
Produktiv use of developer platform build on Cloud Workstations

DZ BANK



Industry: Financial Services
Country: Germany

Customer Award

Harissa
Market price risk application build on GKE

ESG Data
ESG Application in our credit processes build on Cloud Run

2020 — 2021 — 2022 — 2023 — 2024

Landing Zone
Productive use

Flake
Custom data lake powered by BigQuery

Developer Platform
Produktiv use of developer platform build on Cloud Workstations

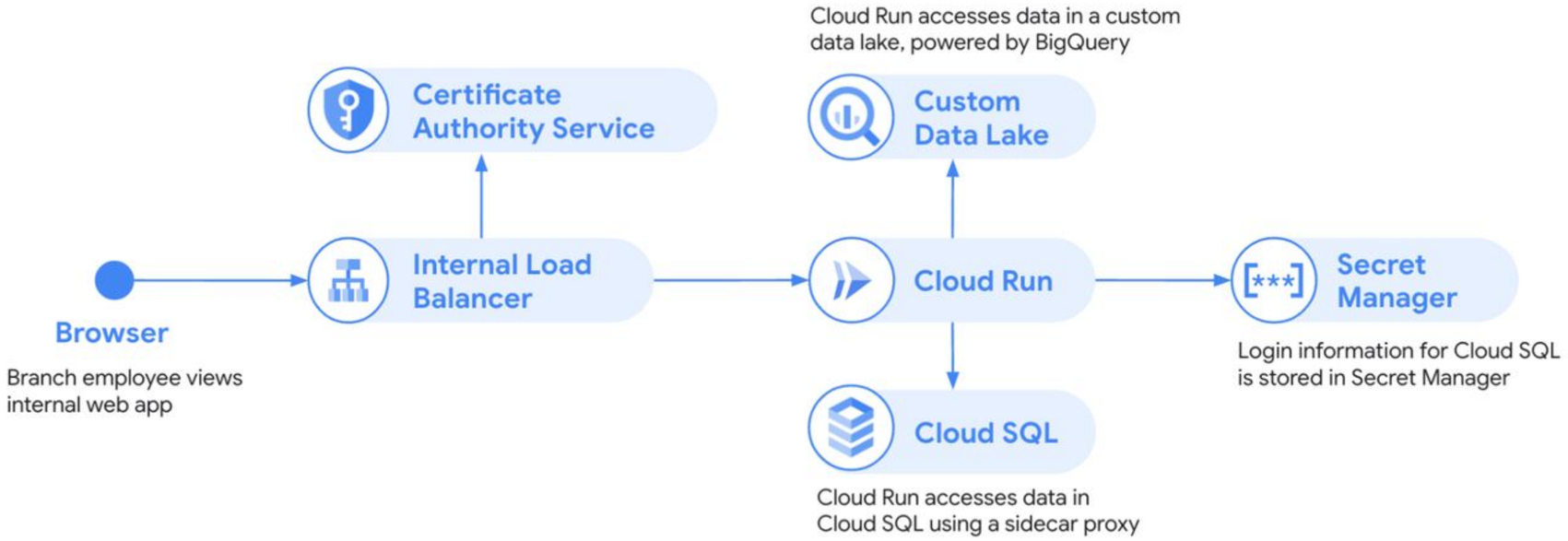
Cloud Run @ DZ BANK
Use Case: ESG Data

Simple architecture

Met rigorous security and compliance standards

70% toil reduction

90% cost savings



What we learned

Safe Gives us an opportunity to innovate and iterate quickly within regulatory and compliance constraints

Easy Automation first with managed services reduces toil and gives our team the gift of time

Efficient Serverless mindset enables us cost savings and easy scalability

In conclusion

Simplifying App Development

Cloud Run continues to focus on new features to simplify the experience of app developers, including Gemini in Cloud Run and the application canvas

Enterprise Ready

With built in security and easy integration with existing systems, Cloud Run can reduce toil and cost of demanding enterprise Workloads

We are interested in your feedback!

Connect with a Product Manager or UX researcher.

Ready to build what's next?

Tap into **special offers** designed to help you **implement what you learned** at Google Cloud Next.

Scan the code to receive personalized guidance from one of our experts.

Or visit **g.co/next/24offers**

Continue your learning journey!

Sessions
DEV310 – Serverless security like a pro

ARC202 – Design secure enterprise networks for a multi-cloud world

Sessions
SEC211 – Next-generation permissions management: Control identities and privileged access to reduce risk

See more from Accenture at **Booth 401**

Demo in Showcase
AIPD-110 Enhance consumer experiences with generative AI

Thank you