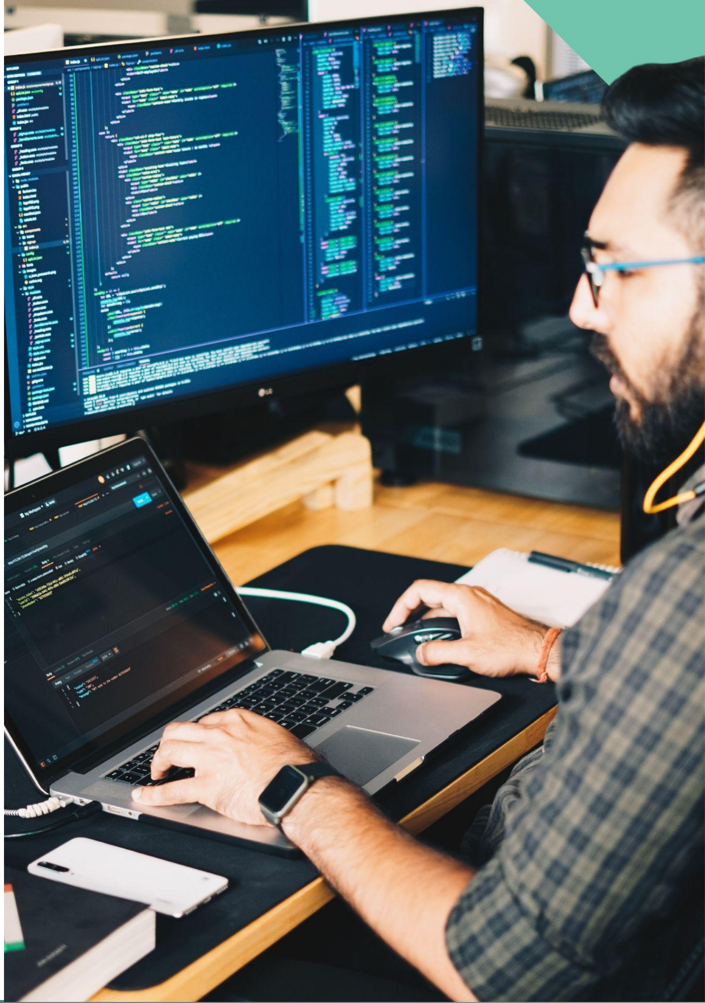
promevo™

Google Cloud

Serverless Workshop: Building Scalable, Secure Applications

October 30, 2023



Copyright © Promevo. Confidential. All rights reserved.

Welcome!

Over the next 45 minutes, we will discuss:

- Real examples of Serverless development
- Serverless Product Manager presentation
- Q&A session with our presenters

promevo™ Google Cloud

Copyright © Promevo. Confidential. All rights reserved.

promevo™

With the expertise, agility, and commitment you can only get from a partner that is solely 100% Google-focused, Promevo is with you every step of the way, enabling your organization to have the best Google life experience possible.

We **Sell**, We **Service,** and We **Build** Google Products

chromeOS

gPanel® by promevo™

Google Cloud

Google Workspace

- **14-Year Google Partnership**
- **Dedicated** Customer Success Team and **Google-Certified** Technical Support Teams
- Ability to drive **license and GCP consumption discounts**
- Custom IT Solutions across **Application, Cloud, and Data Services**
- **Centralized Billing** for all your Google Products and Services
- Proprietary **Google Workspace management platform**

Partnering to Drive Innovation

BrainStorm

CAMEYO

Copyright © Promevo. Confidential. All rights reserved.

Presenters

Justin Barone

Principal Cloud Solutions Architect, Promevo

Aaron Gutierrez

Practice Director, Data Engineering & Analytics, Promevo

Chandni Sharma

Head of Cloud Customer Engineering, Google

Karolína Netolická

Group Product Manager, Cloud Run, Google

Brandon Velasquez,

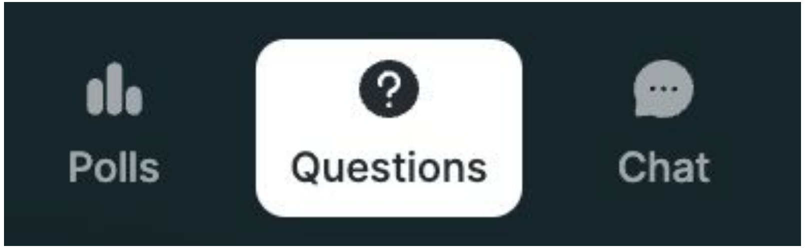
Customer Engineer Google

Daniel Fuentes

Customer Engineer Google

This is Live - Let's Interact!

Questions and Chat



Your Google Cloud Team

Google Cloud

Sales Rep

Lead account strategy, pricing, and overall customer experience

Introduce CE during technical evaluation

Stay in the loop with customer and CE progression

CE

Drive pre-sales technical activities, such as architecture review

Support existing workloads and improve customer experience

Remove technical blockers from customer opportunities

Google Partner

Responsible for implementing, driving migrations, and delivering on solutions and workloads

Collaborate with Sales Reps and CE's to support customer and remove technical blockers

Product

Engage with customers to understand demands and areas of improvement for Google Cloud services

Serve as the voice of the customer within Google Cloud

Copyright © Promevo. Confidential. All rights reserved.

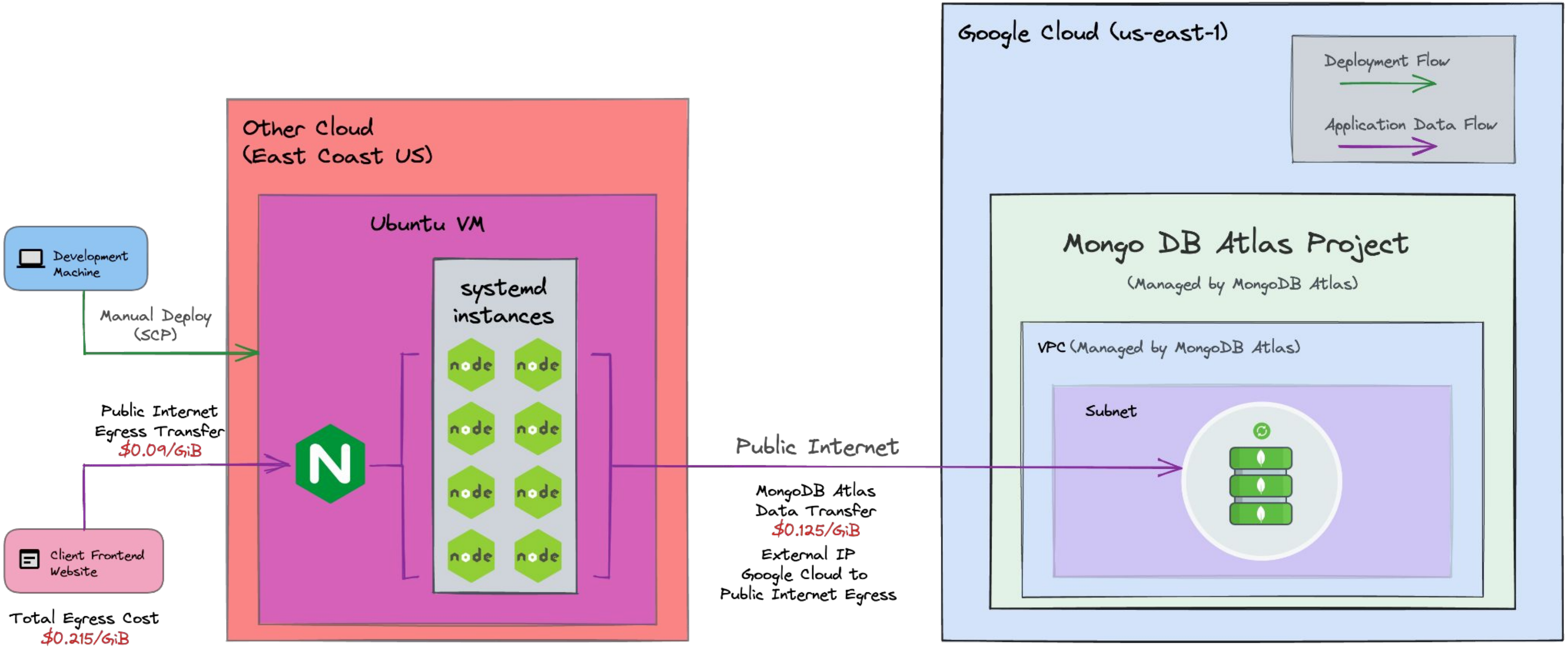
promevo™

Serverless Customer Case Study

The Freedom to Scale: A Cloud Run Success Story

Case Study - Customer Story

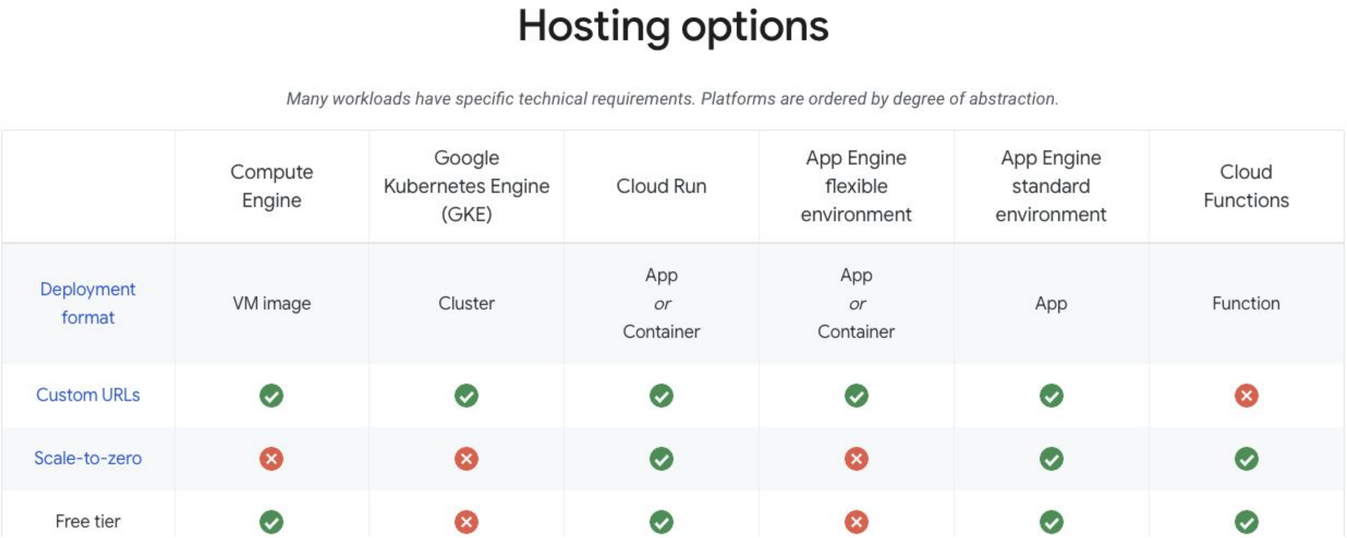
promevo™



Copyright © Promevo. Confidential. All rights reserved.

Which Solution Fit Best? Cloud Run!

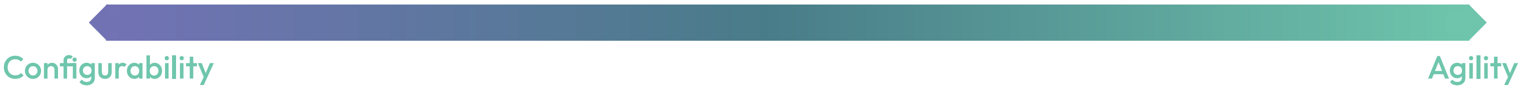
Reference: https://cloud.google.com/hosting-options/



Hosting options

Many workloads have specific technical requirements. Platforms are ordered by degree of abstraction.

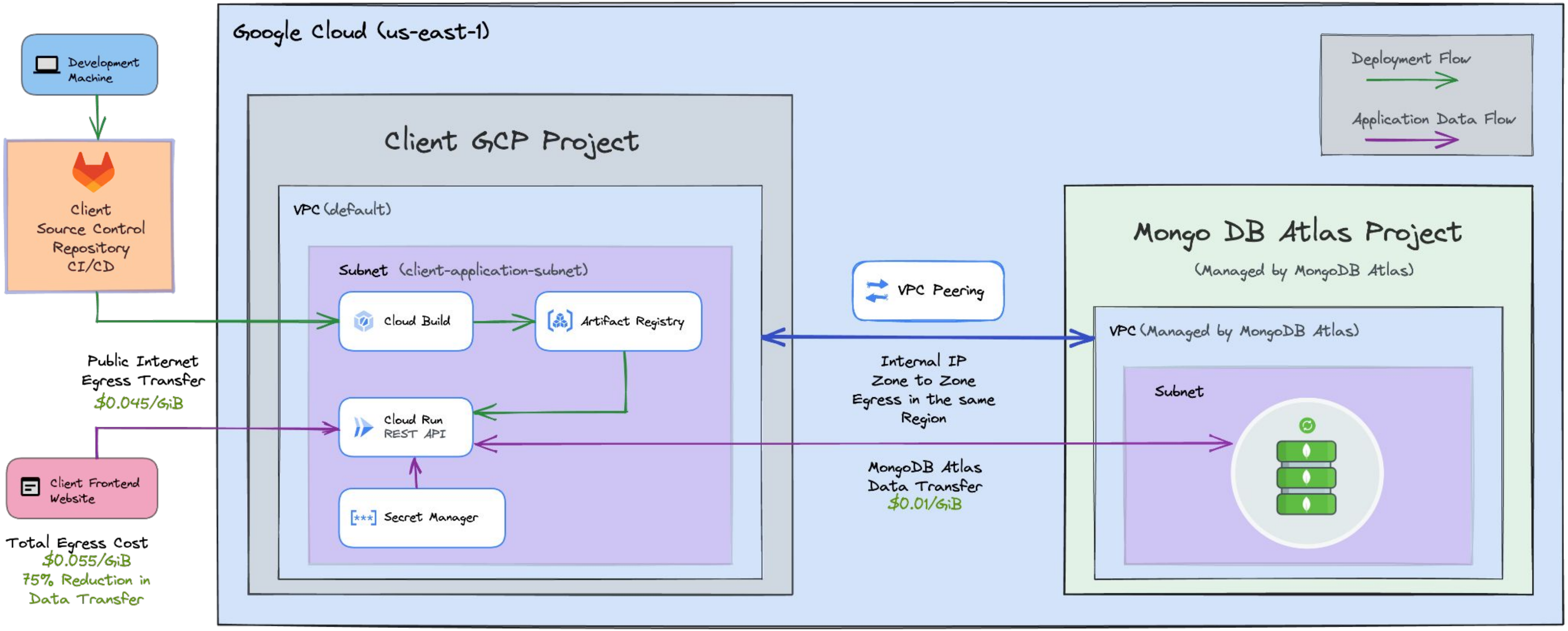
	Compute Engine	Google Kubernetes Engine (GKE)	Cloud Run	App Engine flexible environment	App Engine standard environment	Cloud Functions
Deployment format	VM image	Cluster	App *or* Container	App *or* Container	App	Function
Custom URLs	✓	✓	✓	✓	✓	✗
Scale-to-zero	✗	✗	✓	✗	✓	✓
Free tier	✓	✗	✓	✗	✓	✓



Copyright © Promevo. Confidential. All rights reserved.

Case Study - Customer Story

promevo™



Copyright © Promevo. Confidential. All rights reserved.

How did the Customer Benefit?

Single VM

- Peak 10 req/sec
- Everything manual (Deploys/Scaling)
  Too much human interaction
- Big Egress Costs
- Big Server Costs
- DB although whitelisted was open to the public internet
- 2-3 hours to test & deploy new app versions
- Deploys required maint window
- 22% deployment failure rate
  Human error
- No Monitoring, Alerting, or Metrics
- Zero Redundancy
- Scaling during peak season was a nightmare that had everyone stressed

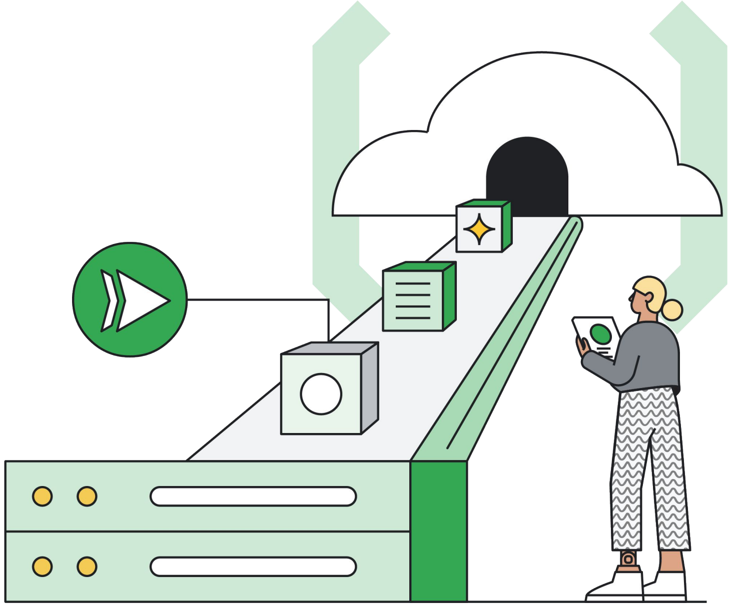
Cloud Run

- Grew to 100 req/sec (no intervention)
- CI/CD Deployment Process
- 75% savings on egress
- DB is more secure
- Dynamically Scale from Zero
- Faster more consistent performance (GCP Premium Network)
- CI/CD test & deploy in less than 4 minutes
  The only bottleneck is automated tests & docker build
- No more maint windows. Deploy multiple times a day.
  Thank you Traffic Splitting
- 1% deployment failure rate
  Still human error
- ROI: Achieved a 98% reduction in deployment time and a 95% decrease in failure rate, yielding an overall efficiency gain of approximately 97%
- DevOps satisfaction 100%

Copyright © Promevo. Confidential. All rights reserved.

Google Cloud

Cloud Run

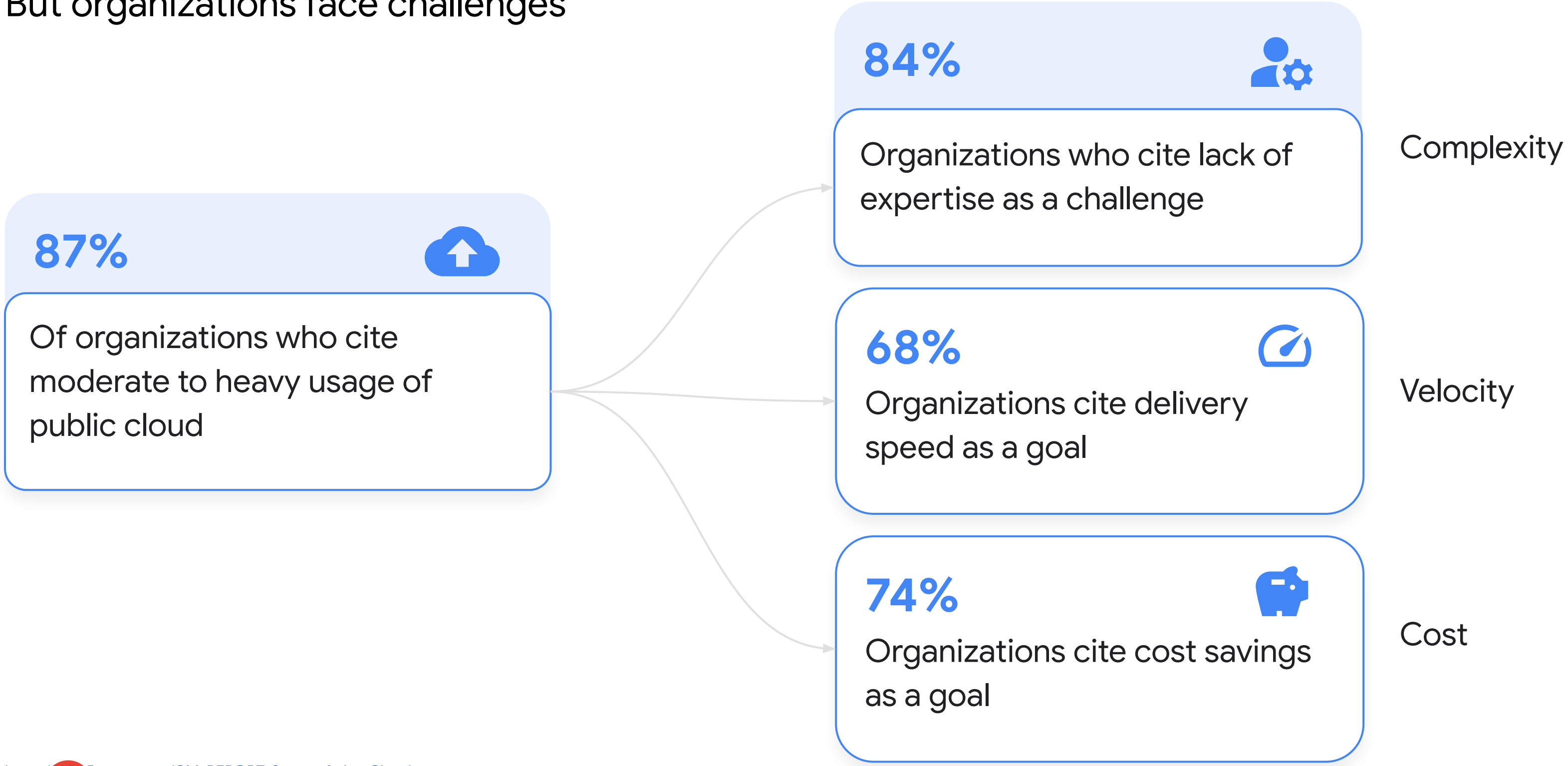


Cloud Functions

Cloud Run

The Future is Cloud

But organizations face challenges



87%

Of organizations who cite moderate to heavy usage of public cloud

84%

Organizations who cite lack of expertise as a challenge

Complexity

68%

Organizations cite delivery speed as a goal

Velocity

74%

Organizations cite cost savings as a goal

Cost

https://info.flexera.com/CM-REPORT-State-of-the-Cloud

Cloud Run

Deploy and scale applications fast and securely in a fully managed environment

1

Simple and automated

Optimized for Developer Velocity

2

Secure

Smaller surface to manage

3

Versatile

Supports many workload types

Simple

Two main resources

Services

Automatically scaled request-driven services

- Out-of-the-box URL with TLS
- Built-in traffic splitting for gradual rollouts
- Can be triggered by events, websockets, HTTP/2 & gRPC
- Pay per request, or per instance lifetime

Jobs

Set of containers which "run to completion"

- Run for up to 24 hours
- No requirement for HTTP
- Runs a specified number of tasks (instances)
- Executed manually, or on a schedule
- Pay only while the job is executing

Demo

Easy to get started

Set up source deployments in just a few clicks.

Easily roll out and roll back revisions.

Easy to operate

Automatically scales in response to traffic.

No pre-provisioning or over-provisioning.

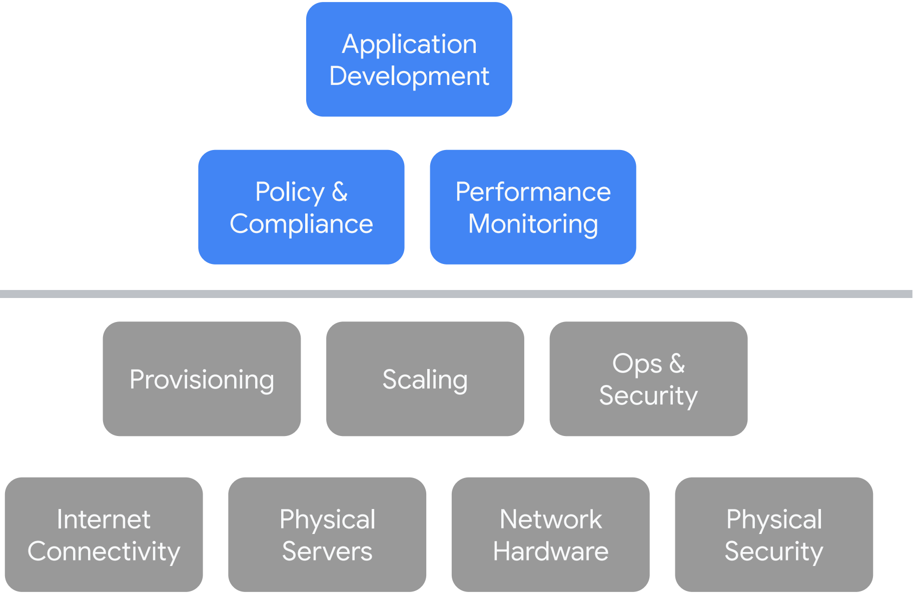
Pay only for what you use

Scales to zero when not in use.

Cloud Run: Simplicity & Velocity

Cloud Run has been designed to **make developers productive** and let them **focus on solving business problems**,

while Cloud Run takes care of the infrastructure.



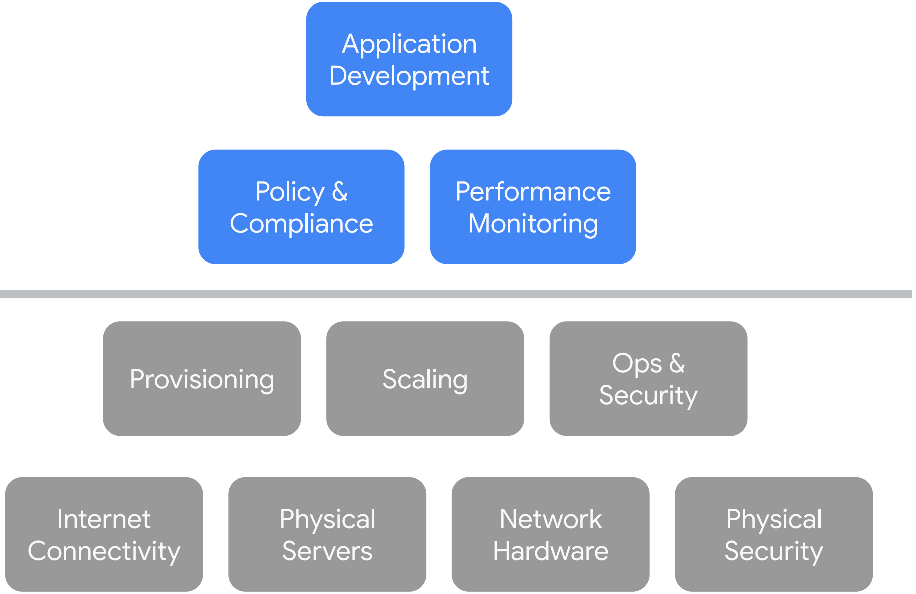
Secure

Smaller product surface = fewer security settings to worry about.

- Your source-to-prod pipeline
- Access controls

Google's responsibilities:

- Container isolation
- Data encryption
- OS patches
- Physical security
- ...



Powerful security features

Your source-do-prod pipeline:

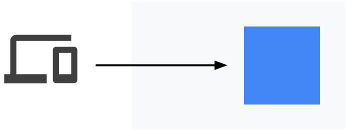
- Scan for vulnerabilities using **Artifact Registry**
- Prevent software supply chain attacks with **Binary Authorization**
- Protect passwords using **Secret Manager**

Access controls:

- Protect services against unauthorized access with **identity-based and network-based access controls**

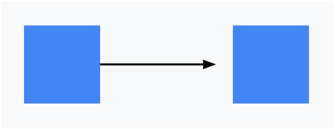
Versatile

Use Cases



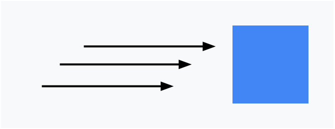
Public Website / API

- Server-side rendered pages
- REST or GraphQL API
- Streaming with WebSockets



Private services

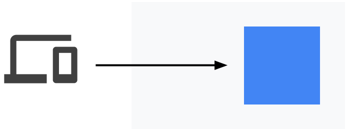
- Internal website or API
- Private HTTP or gRPC microservices



Data processing

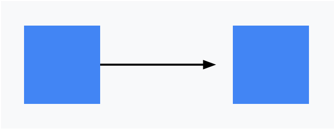
- Process queue messages
- Event driven architecture
- Scheduled Scripts
- Background processing
- Batch Data processing

Use Cases



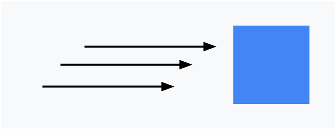
Public Website / API

- Server-side rendered pages
- REST or GraphQL API
- Streaming with WebSockets



Private services

- Internal website or API
- Private HTTP or gRPC microservices



Data processing

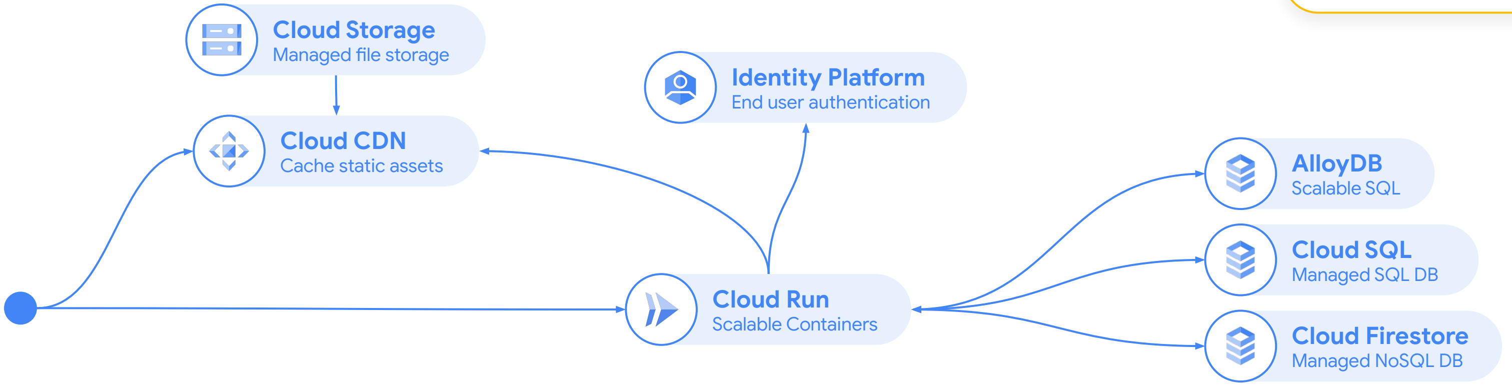
- Process queue messages
- Event driven architecture
- Scheduled Scripts
- Background processing
- Batch Data processing

Design Patterns

Regional Web Application

Serverless Advantage

Containers scale automatically in response to web traffic, including scaling to zero



Deliver

Clients can access public resources through **Cloud CDN** for fast, nearby access to static assets. Assets can come directly from Cloud Storage

Serve

Use **Cloud Run** to handle web traffic. Cloud Run will autoscale based on request traffic, and will be idle when there is no traffic. Cloud Run can also cache responses in Cloud CDN. Use the **Identity Platform** to manage user authentication and authorization

Data

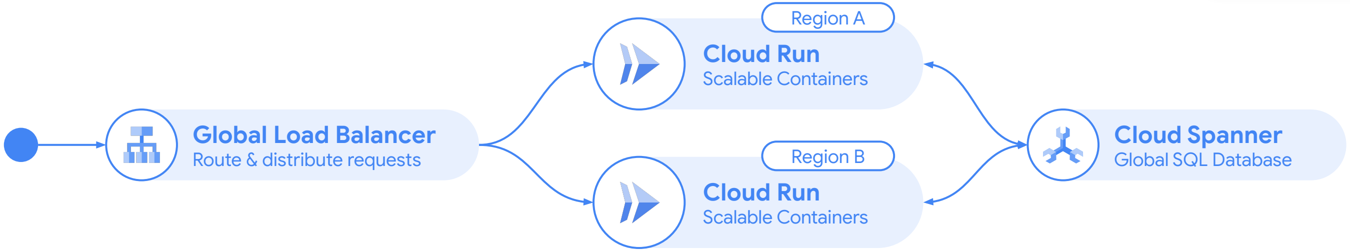
Connect to managed SQL databases like **Cloud SQL, AlloyDB** or managed NoSQL databases like **Cloud Firestore** or **Cloud Bigtable**

Design Patterns

High Availability Web Application or API

Serverless Advantage

Regions with no traffic can scale to zero, so there is minimal incremental cost for each failover region



Deliver

The **Global HTTP Load Balancer** will automatically choose the region closest to the customer, and will route only to available regions

Serve

Cloud Run automatically scales to zero in regions that are not receiving traffic

Data

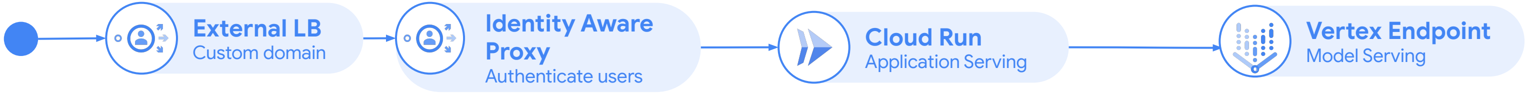
Use **Cloud Spanner** to provide a globally-consistent SQL database with 99.999% availability

Design Patterns

Generative AI application

Serverless Advantage

Using Cloud Run's flexible authentication options is a simple way to manage access to your ML model running in Vertex Endpoints.



Deliver and secure

Use a **Global HTTP Load Balancer** to serve on your own domain, and **Identity Aware Proxy** to authenticate users.

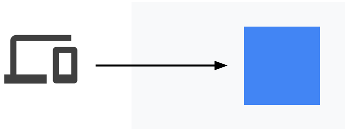
Serve

Serve your application from **Cloud Run,**

Model

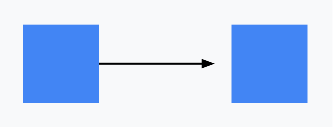
and call your Vertex Endpoint to incorporate AI-generated content.

Use Cases



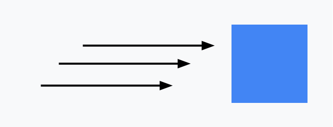
Public Website / API

- Server-side rendered pages
- REST or GraphQL API
- Streaming with WebSockets



Private services

- Internal website or API
- Private HTTP or gRPC microservices



Data processing

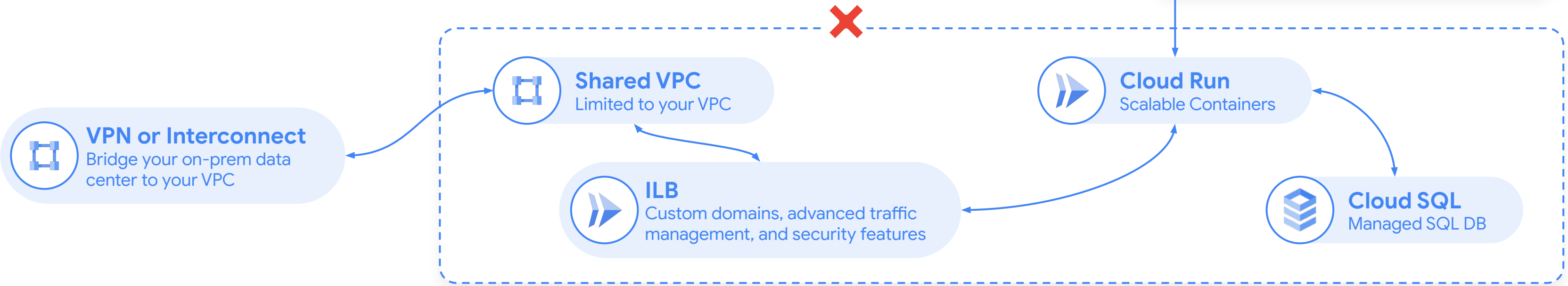
- Process queue messages
- Event driven architecture
- Scheduled Scripts
- Background processing
- Batch Data processing

Design Patterns

Internal Web Services (machine users, not human users)

Serverless Advantage

Lots of internal web services, which don't receive high traffic volume, don't need a fixed infra footprint


VPN or Interconnect
Bridge your on-prem data center to your VPC
Shared VPC
Limited to your VPC
ILB
Custom domains, advanced traffic management, and security features
Cloud Run
Scalable Containers
Cloud SQL
Managed SQL DB

(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

ILB 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

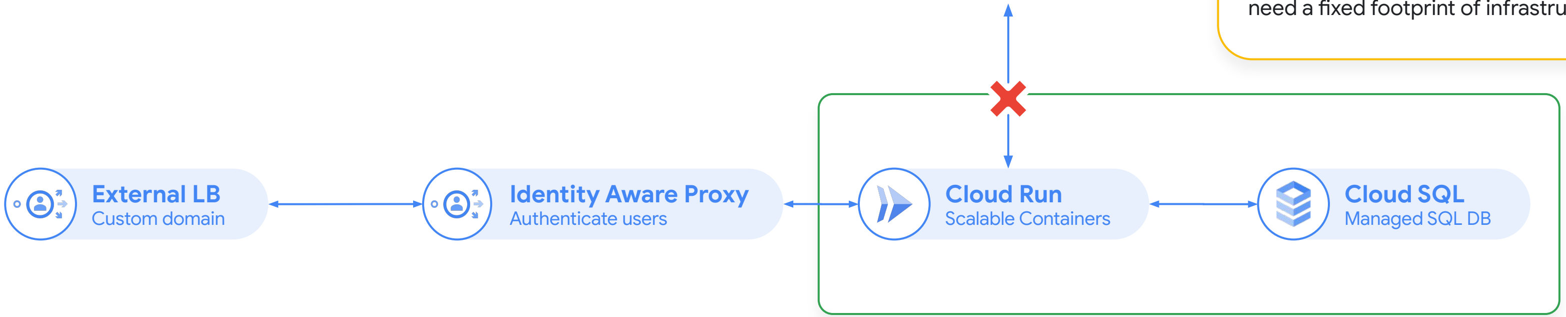
Other Google Cloud resources within the VPC boundary are accessible

Design Patterns

Internal Web Application

Serverless Advantage

Lots of internal apps which don't receive high volumes of traffic don't need a fixed footprint of infrastructure



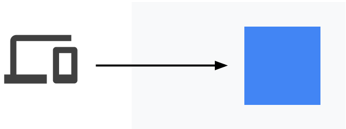
Set up a global **external load balancer** with your custom domain. You can enable additional features for your LB, such as CDN and Cloud Armor.

Authenticate your internal users with **Identity Aware Proxy**

IAP will authenticate to **Cloud Run** so you can ensure that only users that have successfully authenticated to IAP are allowed.

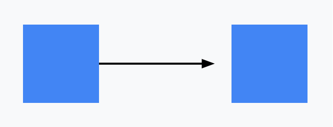
Other Google Cloud resources within the VPC boundary are accessible

Use Cases



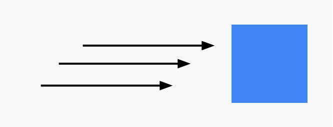
Public Website / API

- Server-side rendered pages
- REST or GraphQL API
- Streaming with WebSockets



Private services

- Internal website or API
- Private HTTP or gRPC microservices



Data processing

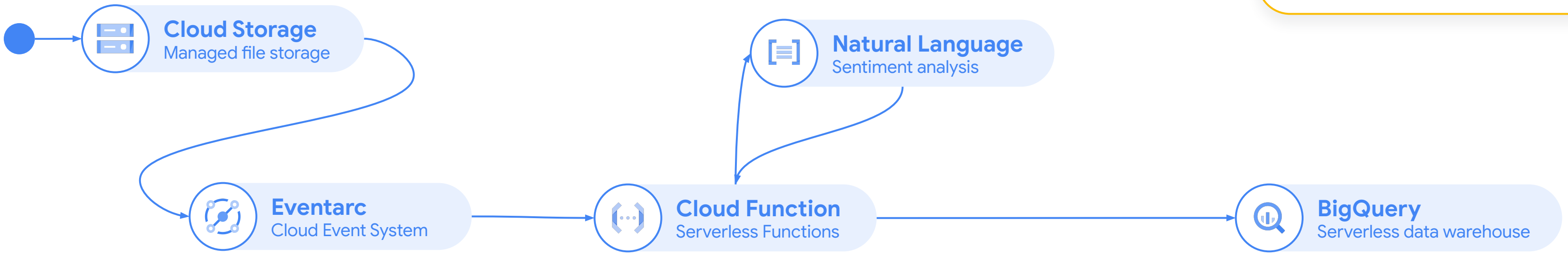
- Process queue messages
- Event driven architecture
- Scheduled Scripts
- Background processing
- Batch Data processing

Design Patterns

On-Demand Data Processing

Serverless Advantage

Easily bind to well-described events and automatically authenticate against other Google Cloud APIs



When a file arrives at Cloud Storage, a **Cloud Event** will be created and handled by **Eventarc**.

A **Cloud Function** is triggered to process the event.

Use the **Natural Language** API analyze the sentiment of text, and use a **Cloud Function** enrich the data.

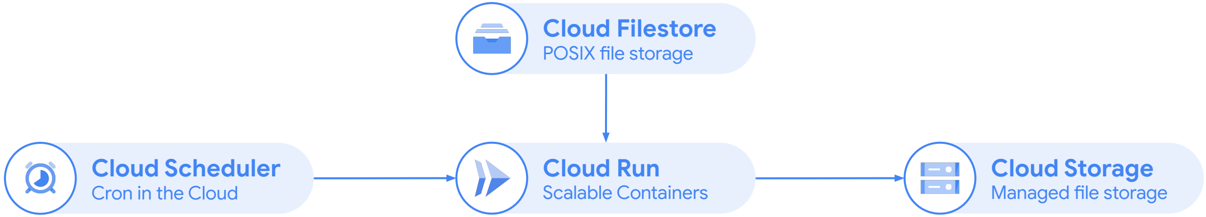
Save the enriched data to **BigQuery**.

Design Patterns

Batch Data Processing

Serverless Advantage

A Cloud Run job can run multiple tasks in parallel, requires no infrastructure setup or provisioning, and scales to zero when complete



Use **Cloud Scheduler** to setup a regular "cron" based on a time/date schema

Use **Cloud Run jobs** to run parallel data processing tasks which run until the container exits (up to 24 hours).

Store processed files in **Cloud Storage**, or any other downstream storage system.

Benefits of Cloud Run

Higher Velocity & Productivity. Serverless allows developers to spend more time writing code and less time managing infrastructure.

Higher Reliability. Serverless is redundant by default. Google is your SRE.

Lower Cost. Serverless autoscales to meet your needs and scales to zero. Pay only for what you use.

95% faster deployment than legacy platforms

98% fewer interruptions to service

15% - 50% cheaper than provisioned platforms

75% cheaper than on-prem

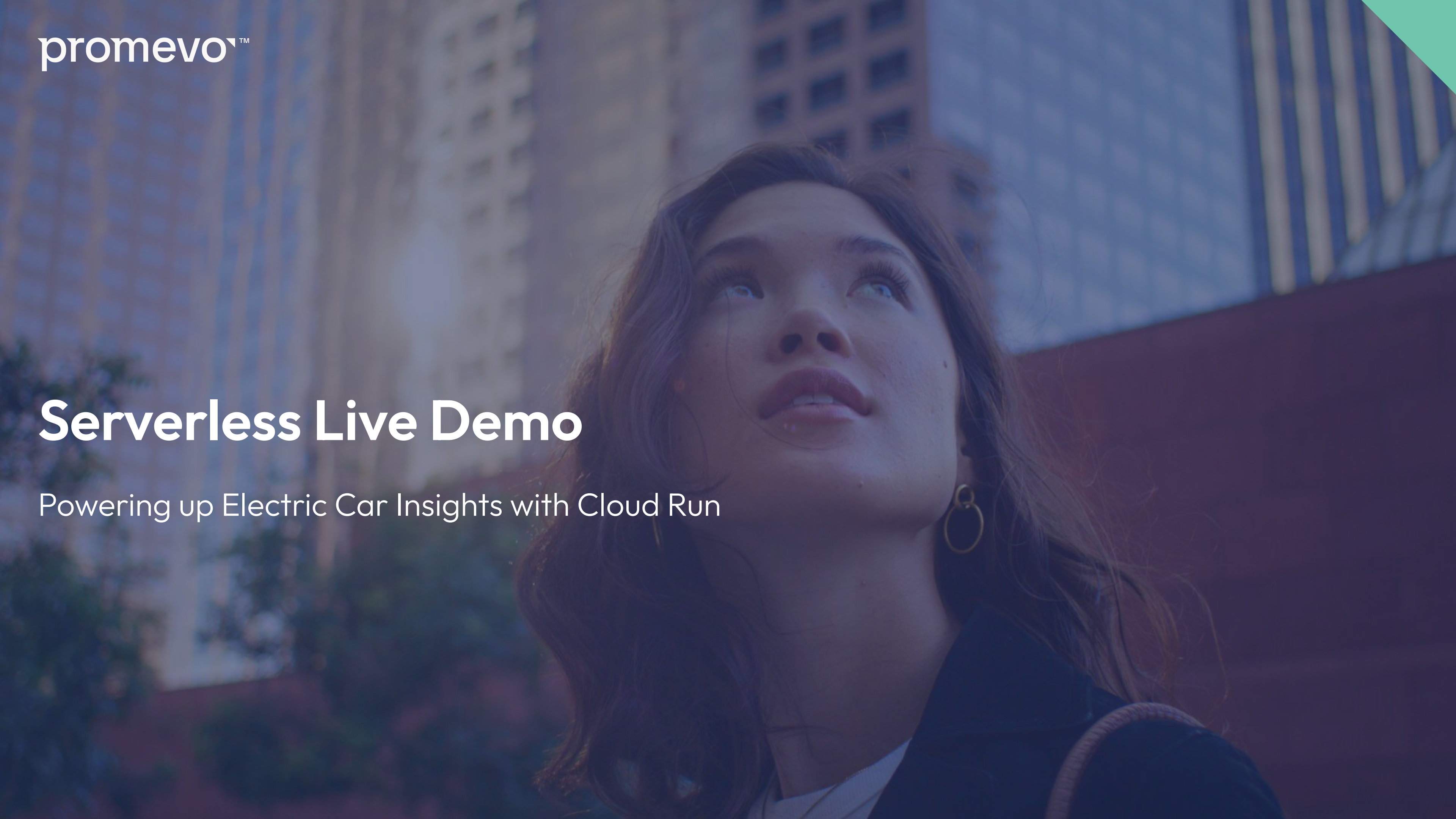
> Our initial concern about choosing serverless was cost.
>
> It turns out that using **Cloud Run is significantly more cost-effective than running the number of VMs** we would need for a system that could survive reasonable traffic spikes with a similar level of confidence.

BBC

Google Cloud

https://cloud.google.com/resources/forrester-cloudrun-benefits-report

promevo™



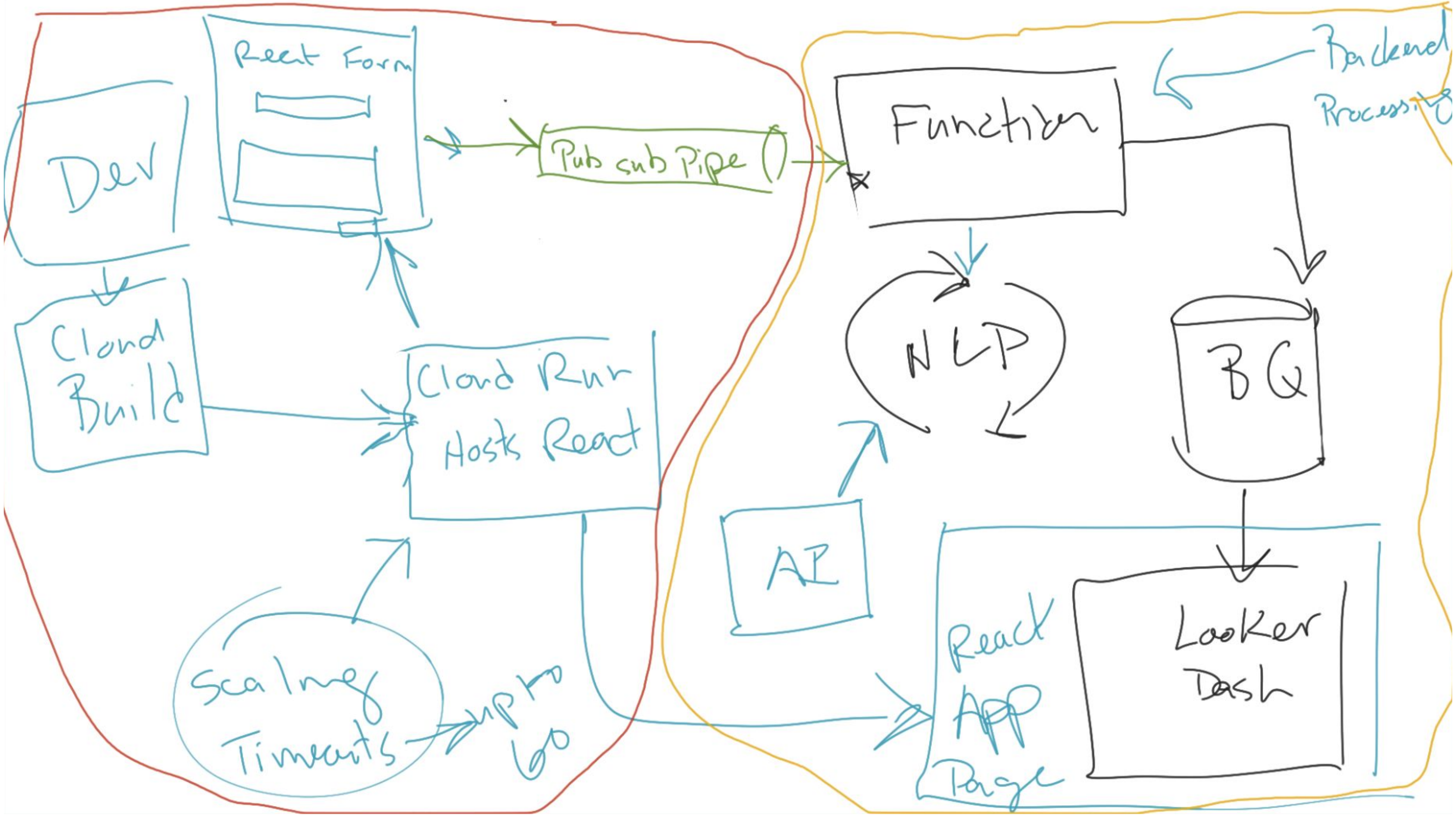
Serverless Live Demo

Powering up Electric Car Insights with Cloud Run

Demo

promevo™

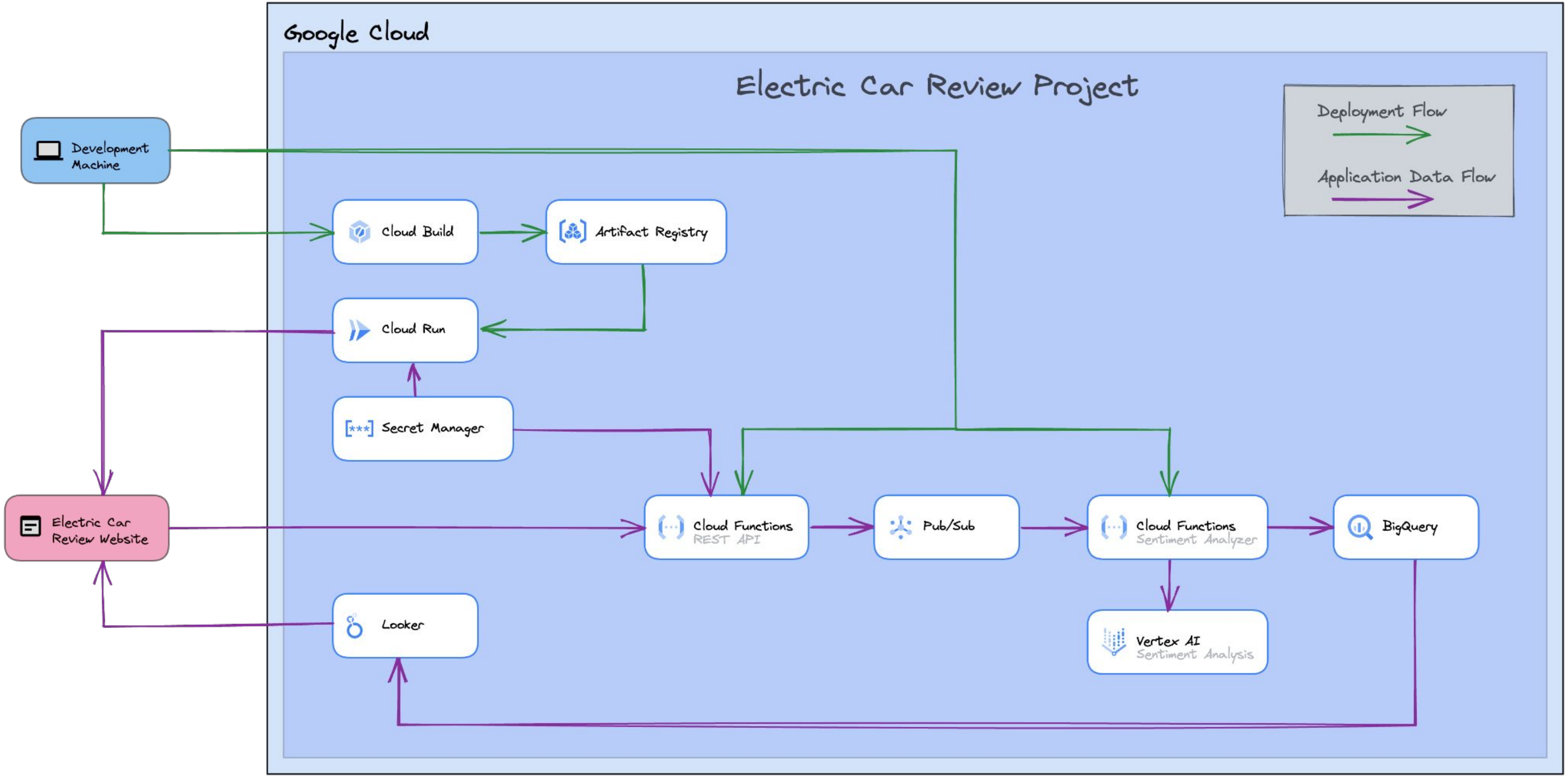
Copyright © Promevo. Confidential. All rights reserved.



Copyright © Promevo. Confidential. All rights reserved.

Architecture Diagram

promevo™

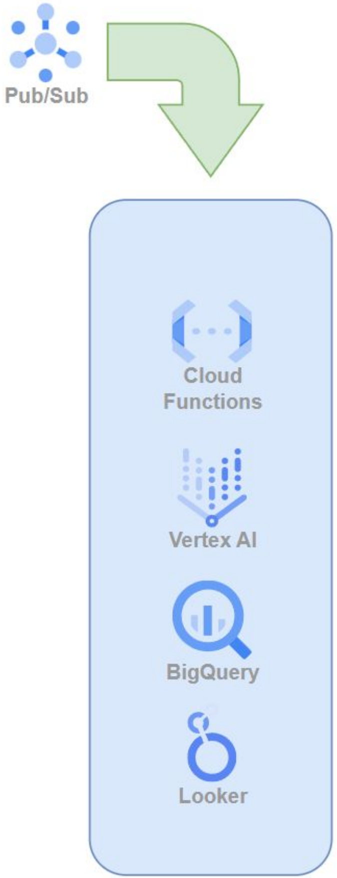


Copyright © Promevo. Confidential. All rights reserved.

Processing The Data

The website will take the user input and send the data through the pipeline. The handoff happens with Pub/Sub. The next steps include the following GCP tools:

- **Cloud Functions**
  Microservice to parse data and orchestrate steps
- **Vertex AI**
  Calls a language model to interpret the user reviews
- **BigQuery**
  Stages and stores the data for use in BI tools
- **Looker**
  Presentation layer

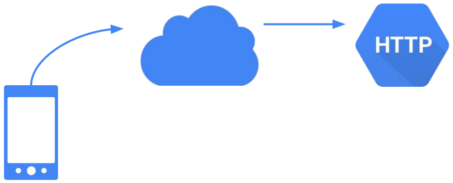


Copyright © Promevo. Confidential. All rights reserved.

Cloud Functions

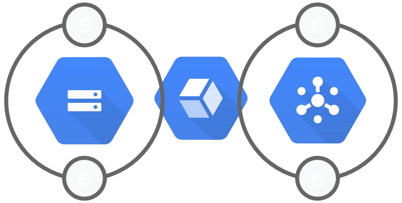
This GCP tool is the best tool for events & async workloads AND/OR single-purpose microservices. Common use cases include:

Webhooks

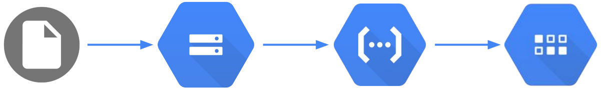


APIs

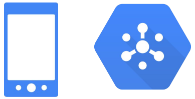
Events



Microservice Backends



Device/iOT analytics

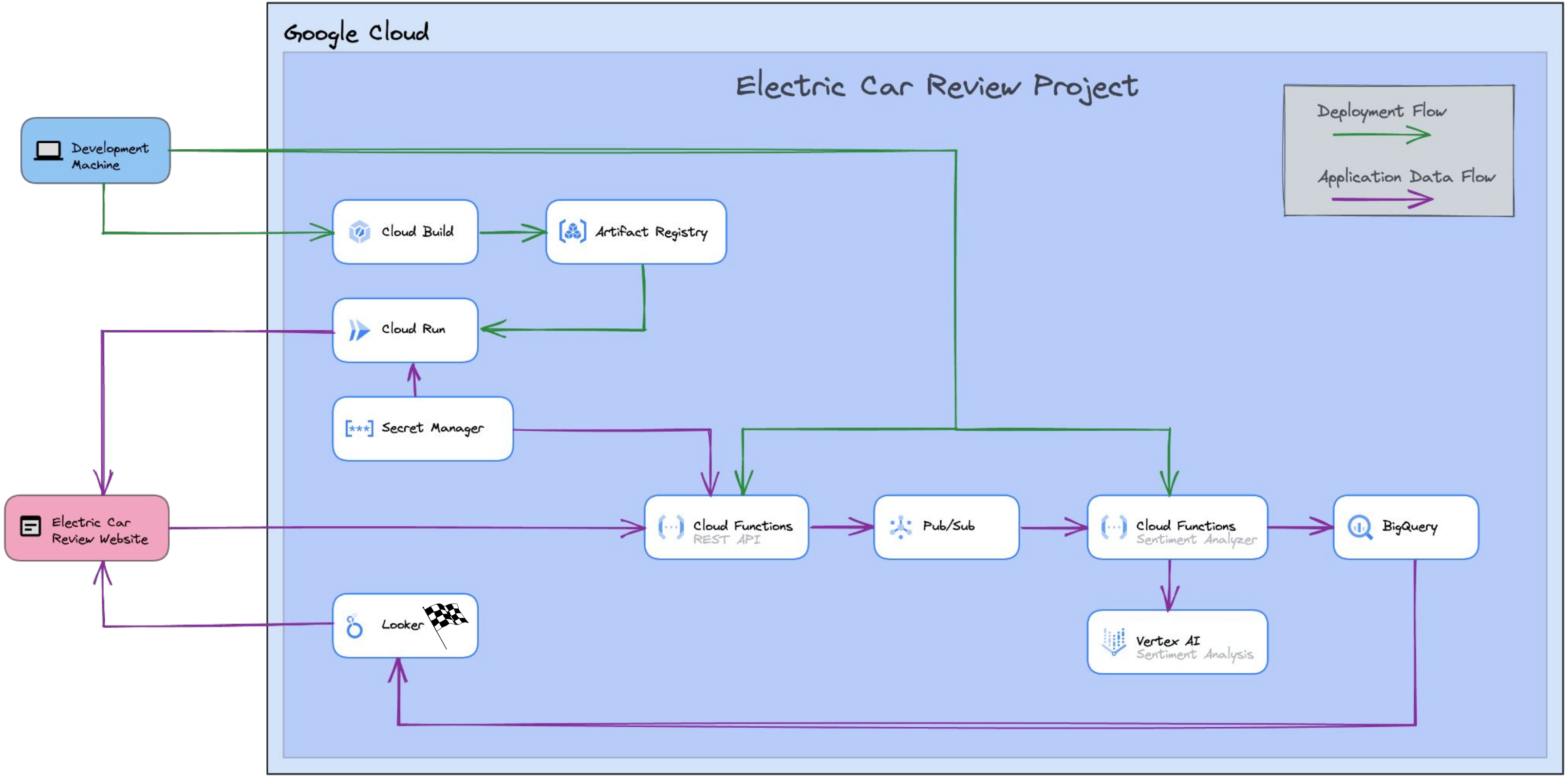


Content transformation/Light ETL

A final note, Cloud Functions have a very generous **Free Tier**!

Copyright © Promevo. Confidential. All rights reserved.

Architecture Diagram



Copyright © Promevo. Confidential. All rights reserved.

Contact Us

promevo.com

promevo.com/gPanel

linkedin.com/company/promevo/

Get ready for our webinar next week by reading our latest blog:

What You Need to Know About Duet AI for Google Workspace

Upcoming Webinars

gPanel® Office Hours
A Promevo Webinar Series

Session 1: The Basics of gPanel

Nov. 14th
Register Here

gPanel® Office Hours
A Promevo Webinar Series

Session 2: Onboarding and User Management

Dec. 5
Register Here

Copyright © Promevo. Confidential. All rights reserved.

Thank you!