



Fundamentals of Cloud Run

Google Cloud

In this module, we discuss the fundamentals of developing and running applications on the Cloud Run platform.

01 Overview

Agenda

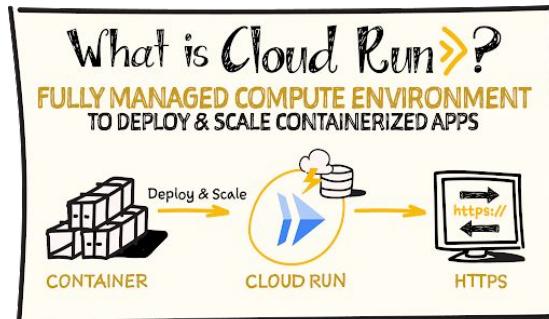


Google Cloud

Let's first discuss what is Cloud Run.

What is Cloud Run?

- A managed compute platform
- Runs containers on Google's infrastructure
- Build full-featured applications with other Google Cloud services



Google Cloud

Cloud Run is a fully managed compute platform that lets you deploy and run containers directly on top of Google's infrastructure.

If you can build a container image of your application code written in any language, you can deploy the application on Cloud Run.

Cloud Run works well with other services on Google Cloud. You can build full-featured applications without spending too much time operating, configuring, and scaling your Cloud Run service.

Services and jobs

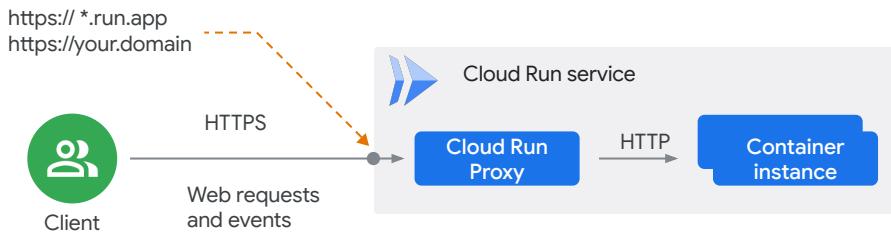


Google Cloud

On Cloud Run, your code can either run continuously as a *service* or as a *job*. Both services and jobs run in the same environment and can use the same integrations with other services on Google Cloud.

- Cloud Run services are used to run code that responds to web requests, or events.
- Cloud Run jobs are used to run code that performs work (a job) and quits when the work is done

Cloud Run supports HTTPS



Google Cloud

A Cloud Run service provides you with the infrastructure required to run a reliable HTTPS endpoint. Your responsibility is to make sure your code listens on a TCP port and handles HTTP requests.

Cloud Run supports secure HTTPS requests to your application, and:

- Provisions a valid TLS certificate, and an HTTPS endpoint to support HTTPS requests. The endpoint is on a unique subdomain of the *.run.app domain. If necessary, you can configure a custom domain for your service.
- Handles incoming requests, decrypts, and forwards them to your application.
- Supports WebSockets, HTTP/2, and gRPC.

Cloud Run jobs



Google Cloud

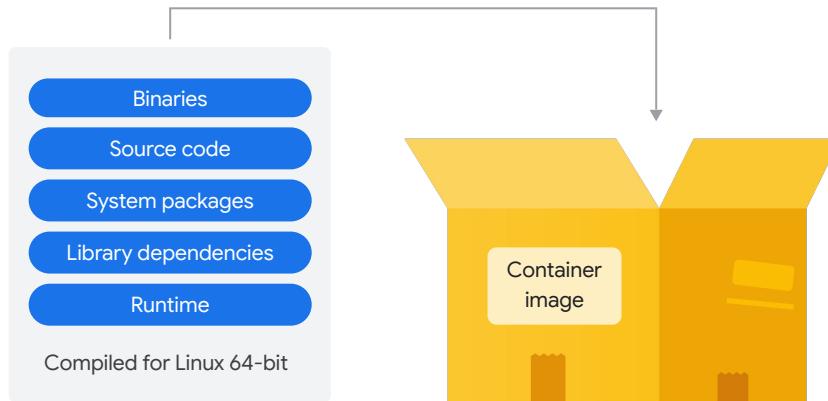
If your code performs work, and then stops (a script is a good example), you can use Cloud Run to run your code as a job.

You can execute a job from the command line using the gcloud CLI, schedule a [recurring job](#), or run it as part of a [workflow](#).

A job can start a single container instance or multiple container instances to run your application code or job script. With multiple container instances running in parallel, the job can complete the task faster. Jobs that run multiple identical container instances are known as Array jobs.

For example, you can use an Array job to process multiple image files from Cloud Storage at the same time with multiple container instances.

Running containers



Google Cloud

For your service or job to be deployable to Cloud Run, you must package it in a container image.

Running containers is a major advantage of Cloud Run. This means that you can develop your applications in any programming language and run them on Cloud Run, as long as they can be compiled to a 64-bit Linux binary and packaged in a container image.

We've discussed containers and container images in the Developing Containerized Applications on Google Cloud course.

Cloud Run features and benefits

Features:

-  Unique HTTPS endpoint for each service
-  Fast request-based auto scaling
-  Built-in traffic management
-  Private and public services
-  Access resources in the VPC network

Benefits:

-  Integrates with Google Cloud services
-  Serverless
-  Supports continuous delivery of code
-  Automatic logging and error reporting
-  Pay-per-use pricing

Google Cloud

Cloud Run supports various features which include:

- Every Cloud Run service is provided with an HTTPS endpoint on a unique subdomain of the [*.run.app](#) domain. You can configure custom domains as well.
- Cloud Run is built to rapidly scale out to handle all incoming requests. A service can rapidly scale up to 1000 container instances. If demand decreases, Cloud Run removes idle containers.
- Every deployment creates a new immutable revision. You can route incoming traffic to the latest revision or roll back to a previous revision. To perform a gradual rollout, you can also split traffic between multiple revisions at the same time.
- A Cloud Run service can be reachable from the internet, or you can restrict access. We discuss access control later in this course.
- Cloud Run container instances can reach resources in the Virtual Private Cloud (VPC) network [through the Serverless VPC Access connector](#).

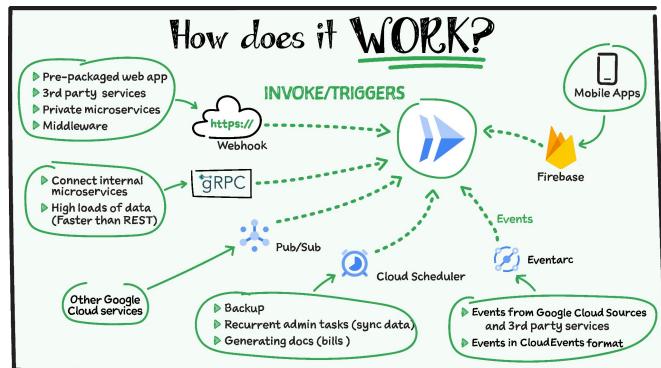
Some benefits of using Cloud Run are:

- Cloud Run integrates with the broader ecosystem of Google Cloud, which enables you to build full-featured applications. You can use:
 - Data storage services such as Cloud SQL, Cloud Storage, Firestore, and others.
 - Cloud Logging for log ingestion and error reporting.

- Identity and Access Management (IAM) for service identification and authentication.
- Cloud Run is serverless so you don't have to worry about infrastructure provisioning and management.
- Cloud Run supports continuous integration and delivery with source code repositories such as GitHub, Bitbucket, or Cloud Source Repositories.
- Pay-per-use pricing for services.

Invoking Cloud Run

- Send HTTPS requests to trigger a Cloud Run service.
- Use gRPC to connect Cloud Run services with other services.
- With WebSockets
- Push messages from Pub/Sub
- Use Cloud Scheduler to run a Cloud Run service on a schedule.
- Use Cloud Tasks
- Trigger a Cloud Run service with events from Eventarc.



Google Cloud

A Cloud Run service can be invoked in the following ways:

HTTPS: You can send HTTPS requests to trigger a Cloud Run service at a stable HTTPS URL. Some use cases include:

- Custom RESTful web API
- Private microservice
- HTTP middleware or reverse proxy for your web applications
- Repackaged web application

gRPC: You can use gRPC to connect Cloud Run services with other services—for example, to provide simple, high-performance communication between internal microservices. gRPC is a good option when you:

- Want to communicate between internal microservices.
- Support high data loads (gRPC uses protocol buffers, which are up to seven times faster than REST calls).
- Need only a simple service definition and you don't want to write a full client library.
- Use streaming gRPCs in your gRPC server to build more responsive applications and APIs.

WebSockets: WebSockets applications are supported on Cloud Run with no additional configuration required.

Trigger from Pub/Sub: You can use Pub/Sub to push messages to the endpoint of

your Cloud Run service, where the messages are then delivered to containers as HTTP requests. Possible use cases include:

- Transforming data after receiving an event upon a file upload to a Cloud Storage bucket.
- Processing your Google Cloud operations suite logs with Cloud Run by exporting them to Pub/Sub.
- Publishing and processing your own custom events from your Cloud Run services.

Running services on a schedule: You can use Cloud Scheduler to securely trigger a Cloud Run service on a schedule. This is similar to using cron jobs. Use cases include:

- Performing backups regularly.
- Performing recurrent administration tasks, such as regenerating a sitemap or deleting old data, content, configurations, synchronizations, or revisions.
- Generating bills or other documents.

Executing asynchronous tasks: You can use [Cloud Tasks](#) to securely schedule a task to be asynchronously processed by a Cloud Run service.

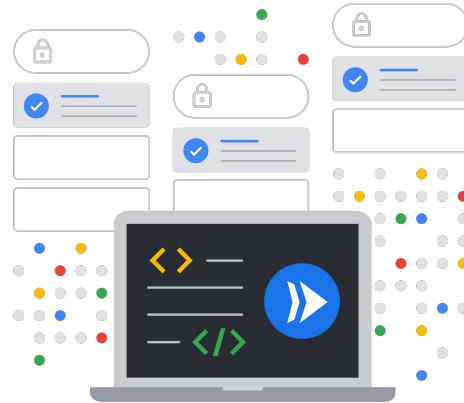
Events from [Eventarc](#): You can trigger Cloud Run with events from various Google Cloud sources.

For example, you can:

- Use a Cloud Storage event (through Cloud Audit Logs) to trigger a data processing pipeline.
- Use a BigQuery event (through Cloud Audit Logs) to initiate downstream processing in Cloud Run each time a job is completed.

Remember

- 1 Cloud Run runs and autoscales containers on-demand.
- 2 Your container-based application handles web requests.
- 3 Cloud Run handles the serving of HTTPS requests to your application.
- 4 Run your application on Cloud Run as a service or as a job.



Google Cloud

In summary:

- Cloud Run is a managed serverless product on Google Cloud that runs and autoscales containers on-demand.
- You can deploy any containerized application that handles web requests.
- Cloud Run handles HTTPS requests to your application.
- Cloud Run runs your application as a service or as a job.