Introduction

Custom container images

Custom containers

Private registries

GPUs and other resources

GPU acceleration

Introduction to Modal

Modal lets you run code in the cloud without having to think about infrastructure.

Features

- Run any code remotely within seconds.
- Define container environments in code (or use one of our pre-built backends).
- Scale up horizontally to thousands of containers.
- Deploy and monitor persistent cron jobs.
- Attach GPUs with a single line of code.
- Serve your functions as web endpoints.
- Use powerful primitives like distributed dictionaries and queues.

Getting started

The nicest thing about all of this is that you don't have to set up any infrastructure. Just:

- 1. Create an account at modal.com
- 2. Install the modal Python package
- 3. Set up a token

...and you can start running jobs right away.

Modal is currently Python-only, but we may support other languages in the future.

How does it work?

Modal takes your code, puts it in a container, and executes it in the cloud.

Where does it run? Modal runs it in its own cloud environment. The benefit is that we solve all the hard infrastructure problems for you, so you don't have to do anything. You don't need to mess with Kubernetes, Docker or even an AWS account.



© 2024

About Status Changelog Documentation

Slack Community Pricing Examples