



NVIDIA's Brev Platform

Develop with the latest NVIDIA GPUs and Software

Introduction to Brev

Your starting point for AI development and experimentation

NVIDIA's Brev platform simplifies AI development by aggregating multiple cloud providers to deliver the right NVIDIA GPUs and environments. Developers can quickly launch and customize GPU-backed setups without complex infrastructure management.

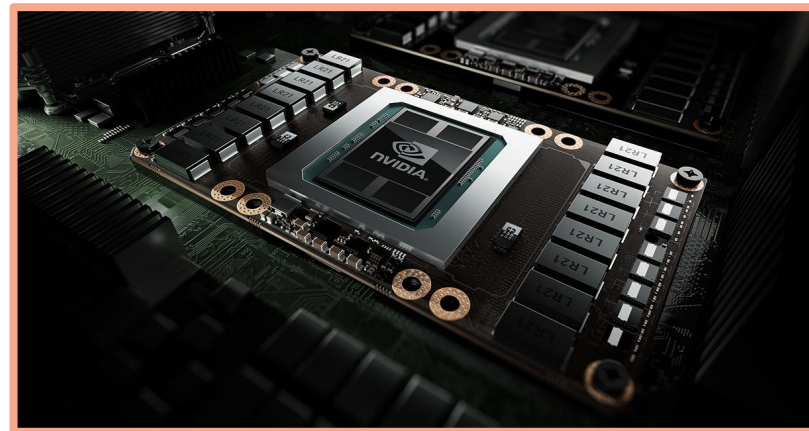
- **Instant GPU access:** On-demand, high-performance NVIDIA GPUs provisioned instantly through our Partners. Virtual Machines preconfigured with *Python*, *CUDA*, *Docker*, and *Jupyter Notebooks*
- **Launchables:** One-click, customizable GPU environments for faster onboarding and team consistency across projects
- **CLI & IDE Support:** Access via Brev CLI and integrate with Cursor, Windsurf, VS Code, and tmux

The screenshot displays the NVIDIA Brev platform interface. At the top, the 'GPU Instances' section includes a description: 'Provision GPU Instances, monitor logs, SSH with the Brev CLI, access JupyterLab, and more!'. A green 'Create New Instance' button is visible. Below this, the 'My Instances' section features tabs for 'All', 'Mine', and 'Team', along with a 'Search Instances' input field. The main workspace area, titled 'BrevGPT Finetune Workspace', shows a workflow for creating a first instance. It includes a 'Compute' section with an 'A100' GPU icon, 'NVIDIA A100 (80GB)', '2 GPUs x 24 CPUs | 340GB', '120GB', and 'GCP' details, with a price of '\$2.48/hr'. This is followed by a 'Container' section with a 'Waiting' status and the image 'nvidia/cuda:12.2-devel-ubuntu22.04'. Finally, a 'Notebook' section shows 'Your FineTune Notebook' with a 'Preview Notebook' link. A 'Create your first instance' link is also present in the workspace header.

Access to Leading Cloud Providers and NVIDIA GPUs

Constantly Growing...

NVIDIA's Brev connects you to a worldwide NVIDIA Cloud Partner network, delivering cutting-edge GPUs on demand for AI/ML acceleration.



Pricing & Packaging

Self-service, pay-as-you-go GPU access for maximum flexibility

- **Flexible Pay-as-You-Go:** Purchase credits to access on-demand GPU compute, ensuring cost-efficiency with no long-term commitments.
- **Transparent Hourly Rates:** Clear pricing per GPU instance type, aggregated from our Partners

Select your Compute

VM Mode w/ Jupyter [Configure](#)

VRAM (GPU Memory) Any Size

Cloud Brev Cloud - any [Filters](#)

B200 **H200** **H100** **A100** **L40S** **A10** **A10G**

1x NVIDIA H200

1x NVIDIA H200 150GiB VRAM 200GiB RAM x 16 CPUS 1TiB SSD SXM NEBIUS

\$4.20/hr

NVIDIA Launchables

“

A one-click experience — no setup, just code running on NVIDIA's Accelerated Computing stack. It's a frictionless way to show, share, and scale innovation.

”

Launchables

1-click deployments for pre-configured software and hardware combinations

The screenshot displays the NVIDIA Launchables interface for a deployment titled "single-cell-analysis-GTC-Paris". The interface includes a top navigation bar with links for Explore, Models, Blueprints, GPUs, Docs, and Forum, along with a Sign in button. The main content area shows the deployment details, including a price of \$1.74/hr, a "Deploy Launchable" button, and a "Content Preview" section. Annotations with green arrows point to specific features: "Price" points to the \$1.74/hr rate; "Predefined Compute" points to the "Compute" section showing the NVIDIA L40S hardware; "Software is pre-configured" points to the "Container" section showing the Docker Compose file; and "Content Preview" points to the "Content Preview" section showing the repository information and README.

Price → \$1.74/hr

Predefined Compute → Compute: NVIDIA L40S (48GiB) | 1 GPUs x 8 CPUs | 147GiB

Software is pre-configured → Container: github.com...docker-compose-nb-2504.yaml

Content Preview → Content Preview: This repository will be present in the instance after deployment

Repository Info

- single-cell-analysis-blueprint by NVIDIA-AI-Blueprints
- 17 stars, 10 forks, 17 watchers

README.md

RAPIDS

Single-Cell Analysis Blueprint

This repository houses tutorial notebooks to run GPU-accelerated single-cell analysis workflows using [RAPIDS-singlecell](#), a GPU accelerated library developed by [scverse](#).

The goal is of this repository is to help users try out and explore different capabilities of RAPIDS-singlecell on datasets ranging from **250 thousand to 11 million cells**. To make this as easy as possible, we set up two different GPU environments on [Brev](#) that are designed to get you working with GPU-



Walkthrough+Demo

- *Create Account*
- *Create Org*
- *Redeem Coupon (if there is one)*
- *Invite Users*
- *2 ways to start:*
 - *1-Click Deploy with Launchables*
 - *Create an instance from scratch*

Happy Coding!

- Brev Docs & Guides: <https://docs.nvidia.com/brev/latest/about-brev.html>
- GPU Catalog: build.nvidia.com/gpu
- Platform or Technical Support: brev-support@nvidia.com