contact@namiq.io · namiq.io

NamiQ

SECURITY ADVISORY

GPUStack exploited

CVE-2026-102895 · on CISA's exploited list since 9 November 2026

Analysis Report

Anyone running GPUStack before version 2.2.0 must update by 12 November 2026 — the CISA deadline. The first step is to confirm which version is installed; if it is below 2.2.0, update immediately, as ...

PREPARED BY	NamiQ	DATE	10 November 2026
FACTS AS OF	10 November 2026	VERSION	1.0
SEVERITY	**Critical: CVSS 3.1 9.8**	STATUS	On CISA's exploited list; due 12 November 2026

NamiQ © 2026 NamiQ. All rights reserved. This document is the property of NamiQ. It may be shared unaltered, with attribution.
Companion article: namiq.io/insights/gpustack-cve-2026-102895

The bottom line

> *Anyone running GPUStack before version 2.2.0 must update by 12 November 2026 — the CISA deadline. The first step is to confirm which version is installed; if it is below 2.2.0, update immediately, as the record says exploitation in the wild has been observed.*

At a glance

Fact	Detail
Product	GPUStack (CVE-2026-102895)
Severity	Critical: 9.8 (CVSS 3.1)
CISA	On its exploited list since 9 November 2026. Deadline for US federal agencies: 12 November 2026.
Chance of exploitation (EPSS)	0.01, the estimated chance in the next 30 days, as at 10 November 2026
Ransomware use	Not recorded by CISA. That is not the same as none.
Fixed in	GPUStack 2.2.0. The record says this issue is fixed in version 2.2.0.
Who looks first	Anyone running GPUStack before version 2.2.0. The record says exploitation was observed in the wild against a 2.1.1 deployment. No account or special condition is needed — network access to the GPUStack API is sufficient.

Does this apply to you?

Answer in order. Stop at the first no.

Question	If no	If yes
Do you run GPUStack anywhere — on-premise, in a VM, or in a container?	Nothing to do for this CVE.	Check which version is running.
Is your GPUStack version below 2.2.0?	No patch needed; confirm the installed version is accurate.	Update to 2.2.0 and check for signs of compromise.
Is the GPUStack server reachable from outside your trusted network?	Treat as medium priority; patch within the CISA window.	Act immediately — the record says exploitation was observed in the wild.

What it is

The flaw lets an unauthenticated remote attacker gain full administrative access by supplying a spoofed loopback address in a request header. The record says they can then change users and models, create persistent admin API keys, read cloud-provider secrets, and run code on the server.

Why it matters

The bypass works by supplying a spoofed loopback address in a request header; no account is needed and the record says this was the default configuration. The record says exploitation was observed in the wild against a 2.1.1 deployment. Once bypassed, an attacker can create persistent administrator API keys, read cloud-provider secrets, and run code through a single API call. Our reading is that any network path to the server — not only the internet — is sufficient for the attack to succeed, because the bypass is in a request header, not in firewall state.

What the vendor's advisory says

We read two pages on github.com. In short:

- The v2.2.0 release notes (18 June) list new features — GPU instance management, Kubernetes Helm chart support, additional database backends — but do not name this CVE or any security fix. The record identifies 2.2.0 as the fixed version.

Things people will get wrong

- The low EPSS score means exploitation is unlikely The record states exploitation was already observed in the wild against version 2.1.1. EPSS reflects statistical modelling, not confirmed activity; confirmed in-the-wild use overrides the score here.
- Only internet-facing instances are at risk The bypass works via a request header any network client can supply. Any host that can reach the GPUStack API, including machines on the same internal network, is sufficient. Internet exposure is not required.

How we would triage it

This is our judgement.

The commands only read: they print a version or list what is installed. They come from general knowledge of this kind of product, not from the sources, so check them against the vendor's page before you run them. Words in capitals are for you to fill in.

Step 1: Find the version you are running

- Check the GPUStack admin interface or run a version query to confirm whether the installed version is below 2.2.0.

```
gpustack --version
pip show gpustack
```

Step 2: Restrict access while you prepare the update

- If the server is internet-facing and cannot be patched immediately, restrict inbound access to the GPUStack API port to trusted IP addresses only.

Step 3: Update to GPUStack 2.2.0

- The record says version 2.2.0 fixes this issue. Follow the vendor's upgrade instructions for your installation method.
- After upgrading, confirm the running version before returning the server to full service.

Step 4: Check for signs of compromise if the server was exposed

- Review the API key list in the admin panel for keys not recognised; remove any suspect keys and rotate all administrator credentials.
- Review server access logs for requests to the inference-backend endpoint and for entries appearing to originate from localhost that arrived from an external address.

To close it, read the version from the product itself, not from an old inventory, and save the output with the date. If you cannot update today, write down who owns it, the new date, and what limits exposure until then. CISA's wording adds that you should stop using the product if no mitigation is available.

Send this to your IT provider

Copy it, change the names and dates, and send it. It is a draft for you to check, not a template tested with any provider.

> *We run GPUStack to manage GPU resources. We have been informed that CVE-2026-102895 (CVSS 9.8) in GPUStack before version 2.2.0 allows unauthenticated administrative access, including code execution. The record says exploitation was observed in the wild. The fix is version 2.2.0. CISA has set a remediation deadline of 12 November 2026. Please confirm: whether GPUStack is installed in our environment, which version is running, and whether version 2.2.0 has been applied or is scheduled. We ask for written confirmation of the installed version and the action taken by 12 November 2026.*

Signs it was used

An update closes the hole. It does not tell you whether someone used it first. CISA's required action points to its forensic triage requirements (BOD 26-04). Look for:

- Requests to the inference-backend endpoint in server logs, particularly those from unexpected source addresses or attributed to localhost from an external interface.
- Administrator API keys in the GPUStack admin panel created at unexpected times or not recognised by any administrator.
- Server log entries showing requests attributed to 127.0.0.1 that arrived on an external network interface.

The weakness behind it

The record classes this as CWE-287 (Improper Authentication) and CWE-290 (Authentication Bypass by Spoofing) and CWE-346 (Origin Validation Error).

Likely ATT&CK mapping

MITRE ATT&CK is a public catalogue of the steps attackers take, from getting in to causing damage. No source has mapped this flaw to it yet, so the steps below are our prediction from what the record says the flaw allows. They are not MITRE's, the vendor's or CISA's, and an attacker may do other things.

Steps an attacker would likely take with this flaw

The step, in plain words	Stage of an attack	Why we think so

Exploit Public-Facing Application (T1190): Breaking in through a service that faces the internet, such as a web server, VPN or login page.	Initial Access: getting in	The flaw is in GPUStack's HTTP API, exploitable over the network with no credentials. This would likely be the initial access technique used to enter the environment.
Account Manipulation (T1098): Changing an existing account, for example adding permissions or keys, to keep access.	Persistence: staying in	The record says an attacker can create persistent administrator API keys after bypassing authentication. This would typically be used to maintain access after the initial bypass.
Command and Scripting Interpreter (T1059): Running commands or scripts on a machine, such as a shell or PowerShell.	Execution: running the attacker's code	The record says code execution is available through the inference-backend endpoint. This would typically involve a command or scripting interpreter on the host.

How to use this: ask your IT provider which control or log would show each step, for example whether sign-in attempts from the internet are recorded. If nobody can answer, that is the gap to close first.

MITRE ATT&CK® is a registered trademark of The MITRE Corporation. Technique names are MITRE's.

The sources, in their words

CISA's catalogue entry:

> *GPUStack contains an improper authentication vulnerability which could allow unauthenticated administrative access, including user and model changes, persistent administrator API-key creation, disclosure of worker and cloud-provider secrets, and code execution through POST /v2/inference-backends/from-yaml.*

The CVE record, published by GitHub on 4 November 2026:

> *GPUStack is an open-source GPU cluster manager for AI model serving and GPU instance provisioning. Prior to 2.2.0, get_current_user in gpustack/api/auth.py treated a loopback request source as the built-in administrator when force_auth_localhost was false, which was the default. A remote request could appear to originate from 127.0.0.1 because client_ip_getter accepted a same-host reverse proxy peer or a client-supplied X-GPUStack-Real-IP loopback value. The bypass allowed unauthenticated administrative access, including user and model changes, persistent administrator API-key creation, disclosure of worker and cloud-provider secrets, and code execution through POST ...*

Versions of gpustack the record lists as affected:

- < 2.2.0

CISA's required action:

> *Apply mitigations in accordance with vendor instructions.*

Check your own project

If you build software, you can compare what a project declares with this CVE from a terminal. The tool reads requirements files, package.json, Maven, Gradle, Cargo, CMake and Dockerfile files, runs nothing, and only the CVE id leaves your machine. Naming the CVEs needs no key.

```
lattice-scan project . --cve CVE-2026-102895
```

Each row says in-range, outside-range, undetermined or lead, with the file and line. That is evidence about what a manifest declares, not about what runs, and a person decides. The verdicts, the CSV columns and a pipeline recipe are in the guide.

Check the servers you run

The same id can be checked against a host. Run this on the machine itself, not on the laptop that deploys to it: it reads installed packages (dpkg, rpm or apk), snaps, the hardware (maker, model, firmware), the running container images with their operating system and packages, and the kernel version, compares them on the spot, and only the CVE ids leave the machine; the inventory stays on it, in ~/.lattice-scan. Naming the CVEs needs no key. Do it on each server after a deployment, and again when a new entry lands.

```
lattice-scan --cve CVE-2026-102895
```

A match is a candidate for a person to confirm, not a finding. A distribution can fix a CVE under an unchanged version number, so a match is a candidate; on Debian-family hosts the report also shows whether the package's own changelog names the CVE. By default it opens the images your containers are running and reads their operating system and package list; a deep scan (--deep, a few minutes) opens every local image and the Python, Node, Java and Ruby packages inside them; for what an image declares, use the project check on its Dockerfile.

Guide: namiq.io/lattice/guides/choose-your-channel. MCP server (io.namiq/lattice in the MCP registry): namiq.io/lattice. Package: pypi.org/project/lattice-scan, or namiq.io/apt.

What we do not know yet

- Whether there is a configuration option that disables the loopback bypass and mitigates the flaw without upgrading, or whether only the version update closes it.
- What specific evidence the in-the-wild exploitation left in logs, and what a compromised instance looks like beyond unexpected API keys.

- Whether cloud-provider credentials that were accessible before patching need to be rotated, and which credential types are at risk.
- We have not tested this against any environment. Versions and scores are as the sources state them and as at 10 November 2026; the vendor's advisory wins where it differs.
- The sections headed "The bottom line", "Does this apply to you?", "What it is", "Why it matters", "Exposure by situation", "Things people will get wrong", "Likely ATT&CK mapping", "How we would triage it" and "Send this to your IT provider" are our interpretation of the sources, not a statement by CISA or the vendor.

Appendix 1. Severity decoded

CVE-2026-102895: score 9.8, `CVSS:3.1/AV:N/AC:L/PR:N/UI:N/S:U/C:H/I:H/A:H` As the CVE record gives it. The record's own score can differ from the base score on the NVD record.

Metric	Value	What it means here
Attack vector	Network	Reachable over a network, with no need to be on the same segment.
Attack complexity	Low	No special conditions beyond the flaw itself.
Privileges required	None	No account is needed.
User interaction	None	No one has to do anything.
Scope	Unchanged	Damage stays inside the vulnerable component.
Confidentiality	High	Total loss of confidentiality.
Integrity	High	Total loss of integrity.
Availability	High	Total loss of availability.

Appendix 2. Timeline

Date	Event
4 November 2026	CVE record published by GitHub.
9 November 2026	CISA added it to its list of exploited vulnerabilities.
12 November 2026	CISA's deadline for US federal agencies.

Appendix 3. Exposure by situation

These priorities are our judgement from the attacker prerequisites in the sources. They are not vendor ratings.

Your situation	Priority	Timeframe
GPUStack before 2.2.0 with its API reachable from the internet	High	Immediately — before 12 November 2026
GPUStack before 2.2.0 on an internal network reachable from untrusted hosts	High	Within the CISA window — by 12 November 2026
GPUStack before 2.2.0 on a fully isolated, trusted internal network	Medium	This week

Running GPUStack 2.2.0 or later	None	No action needed for this CVE

Appendix 4. Definition of done

- Every GPUStack you run is listed, with its version.
- Each is outside the affected list, or the vendor's mitigation is in place.
- The version output or a screenshot is saved with the date.
- Your projects' declared dependencies were checked against this CVE, and every row has a written reason.
- Logs for the weeks before the update were searched, and anything odd was written down.
- Anything not done has an owner and a date.

Sources

- CISA Known Exploited Vulnerabilities catalogue: https://www.cisa.gov/known-exploited-vulnerabilities-catalog
- CVE record for CVE-2026-102895: https://www.cve.org/CVERecord?id=CVE-2026-102895
- FIRST: Exploit Prediction Scoring System: https://www.first.org/epss/
- Advisory or release notes read: github.com/gpustack/gpustack/security/advisories/GHSA-gj9m: https://github.com/gpustack/gpustack/security/advisories/GHSA-gj9m-f8ph-pmhr
- Advisory or release notes read: github.com/gpustack/gpustack/releases/tag/v2.2.0: https://github.com/gpustack/gpustack/releases/tag/v2.2.0
- Reference in the CVE record: github.com: https://github.com/gpustack/gpustack/pull/5346
- Reference in the CVE record: github.com: https://github.com/gpustack/gpustack/commit/53f42a6a553542974396d16a4ff6b12bcbedb48d