Moving to a CDN without overwhelming your origin or blocking Googlebot

22 November 2026 · 6 min read · blogs.iictp.org/cdn-migration-crawl-rate-googlebot-cache-warming

Short answer: after a site moves behind a CDN, crawler traffic to the origin can spike because fast edge responses tell Google's crawlers they can crawl faster, while a cold cache sends every first request for a URL back to the origin. Google's Search Central post of 11 November 2026 advises allowing verified crawlers in your bot rules, pre-warming caches, and using short-lived 429 or 503 responses as a temporary brake.

TL;DR The short version

- After moving to a CDN, crawler traffic to the origin can spike: fast edge responses raise Google's crawl rate while a cold cache sends first requests to the origin.
- Blanket bot protection without exemptions can block Googlebot, so allow verified crawlers in CDN and firewall rules.
- Google's 11 November 2026 advice is to pre-warm caches and use short-lived 429 or 503 responses as a temporary brake.

Read online, listen to it, or change the text size: https://blogs.iictp.org/cdn-migration-crawl-rate-googlebot-cache-warming

What happens when you move to a CDN?

Google's crawling systems calculate a crawl capacity limit for each host. Slow responses lower the crawl rate; fast and stable ones let it rise when there is crawl demand. The moment DNS points at a fast CDN edge, response times drop, and the post says this looks to the crawlers like a hardware upgrade, so they ramp up.

How it works: why the origin gets hit

- A new CDN starts with an empty cache. The first request for each URL at an edge location is a cache miss that goes back to the origin.
- Human visitors cluster on a small set of popular pages that warm up quickly. Crawlers explore the breadth of your URL inventory, so they generate many first-time requests.
- Until the edge caches are populated, the origin is still doing the heavy lifting.
- Google notes that crawlers do not ramp up immediately and many CDNs have connection pacing that smooths spikes.

The common mistake: blanket bot protection

Turning on blanket bot protection without exemptions treats all automated traffic alike. If the CDN or firewall serves a CAPTCHA or challenge page, Google's crawlers cannot pass it. Google says crawlers then back off, and pages, shopping listings and images can begin to drop out of Search.

What the source recommends

- Allow verified bots in CDN and firewall settings; most major CDNs and WAFs have a managed rule for this. Google crawlers can be identified through Google's published IP ranges.
- Better still, use robots.txt or targeted rules for specific unwanted bots instead of blocking all automated clients.
- Pre-warm caches by fetching high-traffic, upcoming-peak and frequently crawled URLs through the CDN at a controlled pace before routing all traffic.
- If load is still critical, return 429 or 503 for a short window, hours or one to two days at most, following Google's guide to urgently reducing crawl rate.

Limitations and open questions

- The post gives no numbers for how much crawl rate rises or how long it takes; it describes the mechanism.
- It concerns Google's crawlers. Other bots and AI crawlers behave differently, so check their documentation.
- Long-running 429 or 503 responses can harm indexing; the post limits the brake to a short window.

What it means for engineers and SEO teams

Treat a CDN launch as a change with a crawl impact. Plan the cache warm-up, agree the verified-bot rule with security, and watch origin logs and Search Console crawl stats for the first days. See what is caching? and what is a load balancer?.

Frequently asked questions

- **Why did crawling go up after I added a CDN?** Faster responses raise the crawl capacity limit, and uncached URLs still reach the origin.
- **Should I block bots at the CDN?** Not blanket-wide. Exempt verified crawlers and target specific bots.
- **How long can I return 503?** Google's post says a few hours, or one to two days at most.

Related reading

- What is caching?
- What is a load balancer?
- What is rate limiting?
- Cloud Architecture Practitioner

Sources and currency

Sources were read on 22 November 2026. Products, rules and standards change, so check the linked originals before you rely on any detail.

Related credential: Cloud Architecture Practitioner. Explore the exam facts and syllabus.

Sources and further reading

1. Google Search Central. Moving to a CDN without overwhelming your origin (or blocking Googlebot). 11 November 2026. developers.google.com/search/blog/2026/11/crawling-cdn.

Published by IICTP, an independent technical credential body, on 22 November 2026. The current version of this post is at https://blogs.iictp.org/cdn-migration-crawl-rate-googlebot-cache-warming. More guides: blogs.iictp.org.