

# Streaming CDN Cheat Sheet

The five tiers, the four dashboards, and the numbers a streaming engineer should know by heart.

## 1. The five-tier streaming CDN hierarchy

| Tier                 | What it does                                     | Typical CHR | Worked load (100k viewers) |
|----------------------|--------------------------------------------------|-------------|----------------------------|
| <b>Origin</b>        | Packager + storage; canonical source             | —           | ~10 rps in                 |
| <b>Origin Shield</b> | Single regional cache funnelling every edge miss | 92-96%      | ~200 rps in                |
| <b>Mid-tier</b>      | Regional cache behind a group of edges           | 85-92%      | ~2,500 rps in              |
| <b>Edge POP</b>      | Metro-area cache, hot content                    | 95-99%      | ~50,000 rps in             |
| <b>Player</b>        | Viewer's device, decode + render                 | —           | 50k viewers                |

## 2. The five jobs of a streaming CDN

| Job                     | Why it matters                                                                                 | Metric to track               |
|-------------------------|------------------------------------------------------------------------------------------------|-------------------------------|
| <b>Shorten distance</b> | Edges in the viewer's metro area cut Time to First Byte from 300-800 ms to under 50 ms.        | <b>TTFB per region</b>        |
| <b>Absorb load</b>      | 100k viewers × 1 seg / 2 s = 50,000 rps; CDN absorbs 95-99%.                                   | <b>Cache hit ratio</b>        |
| <b>Resilience</b>       | Thousands of edges; stale-while-revalidate keeps content flowing during origin outages.        | <b>Per-region uptime</b>      |
| <b>Edge policy</b>      | Signed URLs, token auth, geo-blocks enforced at the edge — microseconds, not round-trips.      | <b>Blocked-request rate</b>   |
| <b>Billing</b>          | One bill replaces dozens of transit / peering / colo contracts. 2026 list: \$0.002-\$0.085/GB. | <b>\$/GB, 95th-percentile</b> |

## 3. Four dashboards to read weekly

- Request volume per edge group — find the hot regions.
- Cache hit ratio per layer — < 80% means misconfigured cache keys.
- Origin egress bytes — should be a small fraction of edge egress.
- Per-region TTFB — pinpoints which geographies need a closer POP or a multi-CDN partner.

### Worked savings — one afternoon of origin shield

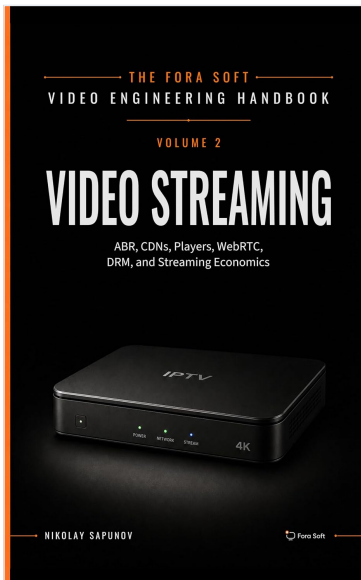
- **50,000 viewers @ 3 Mbps** — regional sports streamer
- **12,500 segment fetches / sec** — 1 seg every 4 s, per viewer
- **Without shield: 4.5 Gbps origin egress** — ~\$180K / month
- **With shield: 0.36 Gbps origin egress** — ~\$14.4K / month
- **12.5× egress reduction** — ≈ \$150K / month saved

## 4. Cache-hit-ratio pitfalls and the one-line fix

| Symptom                                                       | Fix                                                     |
|---------------------------------------------------------------|---------------------------------------------------------|
| <b>Per-user query string in cache key (token, session-id)</b> | Strip from cache key; verify token at edge separately   |
| <b>SSAI inserts ads per viewer in the manifest</b>            | Switch to SGAI — manifest personalised, segments shared |
| <b>VOD Cache-Control: max-age=60</b>                          | Set days or weeks; movies don't change every minute     |
| <b>URL purge on every re-encode</b>                           | Use versioned filenames (segment-v2.m4s) — no purge     |

Companion to: 'What a CDN Actually Is, for the Streaming Engineer'

[www.forasoft.com/learn/video-streaming/articles-streaming/cdn-for-streaming-engineer](http://www.forasoft.com/learn/video-streaming/articles-streaming/cdn-for-streaming-engineer)



GO DEEPER · VOLUME 3 OF 7

## Video Streaming: A Complete Guide to Video Delivery

From ABR and CDNs to Players, WebRTC, DRM, and the Economics of Streaming

By Nikolay Sapunov, CEO at Fora Soft · 1,182 pages · Kindle & paperback on Amazon

[Read it on Amazon →](#)
[amazon.com/dp/B0HH76ZYWC](https://amazon.com/dp/B0HH76ZYWC)

This sheet gives you the five CDN tiers, the four dashboards and the cache-key mistakes that sink hit ratio. Volume 3 covers CDNs and caching in full, along with multi-CDN strategy, ABR ladders that shape edge traffic, QoE monitoring and the economics of delivery.

### WHAT YOU WILL FIND INSIDE

- Cache hierarchy, origin shield and request collapsing
- Multi-CDN switching and per-region QoE
- Cache keys, TTLs and purge strategy for HLS and DASH
- Egress pricing and 95th-percentile billing

### THE FULL HANDBOOK · 7 VOLUMES · 9,000+ PAGES · ALL ON AMAZON

Vol. 1 · Video Encoding — pixels and color to H.264, HEVC, AV1, quality metrics and AI

Vol. 2 · Video Streaming I — network transport, contribution and ingest, HLS, MPEG-DASH, CMAF

**Vol. 3 · Video Streaming II — ABR, CDNs, players, WebRTC, DRM and the economics of delivery**

Vol. 4 · Audio for Video I — PCM, codecs, loudness, accessibility, real-time voice processing

Vol. 5 · Audio for Video II — jitter buffers, lip sync, Atmos, quality metrics, neural codecs

Vol. 6 · AI for Video I — models and inference, computer vision, speech, search, generative video

Vol. 7 · AI for Video II — real-time pipelines, agents, serving, cost, compliance, verticals

## Is your origin egress or CDN bill higher than it should be?

Share your CDN logs and current architecture; we will find the cache-key and shield gaps and estimate the monthly saving before any change ships. 250+ video products shipped since 2005.

[Book a 30-min call →](#)
[WhatsApp →](#)
[Email us →](#)