

CDN Delivery Readiness Checklist

Audit your video delivery before you sign a CDN contract: measure the right number, set TTLs deliberately, keep the cache key clean, and shield the origin. Engineering guidance — confirm vendor numbers live, they change.

1 · MEASURE THE RIGHT NUMBER (offload, not just CHR)

- Track byte-based origin offload, not only request cache-hit ratio.** Large segments dominate the bill, so the bytes you keep off origin — not the count of requests — is the cost metric.
- Target 95%+ offload for popular VOD.** Remember the math is not linear: 90→95% halves origin load, and 95→99% halves it again. The last points are worth the most.
- Write your numbers:** current offload _____ % · target _____ %

2 · SET TTLs DELIBERATELY (segments long, live manifest short)

- Cache immutable segments LONG.** A high s-maxage (RFC 9111) on segments is the main driver of offload — a segment never changes, so the edge can hold it for days.
- VOD manifest: long. LIVE manifest: only ~½ its segment duration.** A stale live manifest hands players an outdated segment list and stalls playback.
- Write your TTLs:** segment _____ · live manifest _____

THE ONE SANITY CHECK BEFORE YOU SIGN

Count bytes, not features. Origin egress = bytes delivered to viewers × (1 – offload ratio). Worked: 100,000 viewers × 10 h × 5 Mbps ≈ 2.25 PB delivered per month. At 90% offload the origin sends 225,000 GB; at 95% it sends 112,500 GB — exactly half. At a representative ~\$0.08/GB origin egress that single step is about \$9,000/month (~\$108,000/year), before the origin compute and capacity you no longer provision for the doubled load. If your offload sits below 95% on popular titles, that is the first cost to cut. Note: the CDN bill on the bytes the edge sends to viewers is a separate, larger line — model it in the CDN egress article.

3 · KEEP THE CACHE KEY CLEAN (one URL for everyone)

- Strip per-viewer query strings** (session IDs, analytics tags, cache-busters) from the cache key. Left in, they fork one popular segment into thousands of per-viewer copies and collapse offload.
- Confirm every viewer of a rendition converges on one cached segment URL.** This is the property that lets one copy serve the crowd.
- Tokenized URLs are fine** — but keyed so they still share the cached object. Your cache-key rule: _____

4 · PROTECT THE ORIGIN (shield before the premiere)

- Turn on an origin shield / tiered cache with request collapsing.** One designated mid-tier cache sends a single origin fetch per object during a stampede, instead of one per edge.
- Verify a global PoP footprint for your actual viewer regions.** Distance is latency; a viewer far from any PoP buffers no matter how high your offload.
- Confirm:** origin shield in place? _____ · regions covered _____