Live-Event Readiness & Premiere-Spike Worksheet

Pressure-test an unrepeatable live premiere before the night. Size BOTH herds, harden the origin, plan failover and degradation, then rehearse at peak. Engineering guidance — CDN egress is tiered, commit-dependent, and 95th-percentile billed, so confirm your own rate and date the assumption.

1 · SIZE BOTH HERDS — fill in for your PEAK (not registrations, not average)

- [ ] **Peak concurrency.** Highest simultaneous streams in the first minute. Size everything to this. _____
- [ ] **Data-plane peak = concurrency × bitrate.** 2,000,000 × 5 Mbps = 10 Tbps. Your figure: _____
- [ ] **Manifest herd = concurrency ÷ (½ × target duration).** 2,000,000 ÷ 3 s ≈ 667k req/s, sustained.
- [ ] **Control burst = (concurrency × calls) ÷ arrival window.** (2,000,000 × 3) ÷ 90 s ≈ 67k req/s on auth + entitlement.
- [ ] **Egress = Tbps ÷ 8 × seconds.** 10 Tbps × 2 h ≈ 9 PB; at $0.02/GB ≈ $180,000 — one show may set the 95th-pct month.

2 · HARDEN THE ORIGIN (tick what your design already does)

- [ ] Edge request collapsing on — many identical fetches become one upstream fetch.
- [ ] Origin shield enabled — all edges collapse to ≈1 origin fetch per segment.
- [ ] Cache-hit / offload ratio targeted above 90% for the live edge.
- [ ] Pre-warm what exists in advance (slate, pre-show, VOD angles); CMCD (CTA-5004) prefetch hints on.

3 · RELIABILITY + GRACEFUL DEGRADATION (an event you cannot re-run)

- [ ] More than one CDN, delivery kept portable; failover rehearsed (not just wired).
- [ ] Content steering (ETSI TS 103 998 / EXT-X-CONTENT-STEERING) — move CDNs without a restart.
- [ ] Admission control (fail-open waiting room) + bitrate shed (cap ladder) + static slate floor.
- [ ] Retry with exponential backoff + jitter — never a fixed timer, or recovery is a second herd.

4 · BEFORE THE BIG NIGHT (the readiness gate)

- [ ] Forecast PEAK concurrency separately for steady catalog and the live spike?
- [ ] Capacity committed AND the CDN notified of date, peak, and regions in advance?
- [ ] Control plane provisioned AHEAD of the arrival curve (predictive, not reactive)?
- [ ] Whole platform load-tested at target peak in REHEARSAL — never first in production?
- [ ] Failover + degradation drills fired, and a real-time monitoring + on-call plan set?

THE ONE RULE OF LIVE

A live premiere is two herds in one minute: a control-plane herd (requests/sec on sign-in, entitlement, the first manifest) and a data-plane herd (terabits/sec as everyone pulls the same fresh segment). Live is the hardest delivery because the edge is always near-cold — the segment everyone wants is two seconds old, so the thundering herd re-forms on every segment for the whole event. Turn a million identical fetches into one with request collapsing and an origin shield, provision ahead of the arrival curve, fail over across more than one CDN, and degrade gracefully instead of breaking. Then rehearse the whole platform at peak before the night — because the event cannot be re-run.



GO DEEPER · THE 7-VOLUME HANDBOOK

Fora Soft Video Engineering Handbook

Seven volumes, 9,000+ pages: encoding, streaming, audio and AI for video products — from the pixel to production.

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

Browse the handbook on Amazon →

This worksheet sizes both herds of a live premiere and gates the night on origin, failover and rehearsal. Our seven-volume handbook goes deeper into that delivery path: start with Volume 3 for CDNs, caching and the economics of delivery, then Volume 2 for ingest, HLS and DASH, and Volume 1 for encoding.

WHAT YOU WILL FIND INSIDE

- CDN caching and multi-CDN delivery (Vol. 3)
- Egress cost and delivery economics (Vol. 3)
- Live ingest, HLS and DASH (Vol. 2)

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

Preparing a live premiere or sports event that cannot fail at peak?

We will size concurrency and egress, review origin shielding and failover, and plan a load test at your target peak. 250+ video products shipped since 2005.

Book a 30-min call → **WhatsApp →** **Email us →**