Streaming Operations & Alerting Runbook

Running an OTT platform live is two jobs: a dashboard that informs and an alert that interrupts. Watch four golden signals, page only on symptoms a viewer feels, size alerts to how fast the error budget burns, and back it all with a sane on-call model. The rule under everything: a pager that fires for everything is a pager nobody reads.

1 · THE FOUR GOLDEN SIGNALS (WATCH THESE)

- [ ] **Latency** — startup time, manifest fetch, segment delivery; track p99 and failed-request latency separately.
- [ ] **Traffic** — concurrent viewers, requests/sec, egress; concurrency is the live-event heartbeat.
- [ ] **Errors** — playback failures, CDN 4xx/5xx, license errors; use the rate, never the raw count.
- [ ] **Saturation** — origin, transcoder, cache-hit, license-server headroom; rising tail latency warns first.

2 · ALERT ON SYMPTOMS, NOT CAUSES

- [ ] **Symptom, not cause** — page on 'playback failing in Germany', not on every internal 503.
- [ ] **Pass the four tests** — urgent, actionable, user-visible, and not already paging someone else.
- [ ] **Right representation** — percentile not average, rate not count, rate-of-change not current.
- [ ] **Kill alert fatigue** — move non-urgent causes to the dashboard; keep only pages worth waking for.

3 · SLO → ERROR BUDGET → BURN RATE

- [ ] **Set the SLO** — e.g. 99.9% of requests succeed; keep the customer SLA looser so you trip first.
- [ ] **Error budget** — 100% − 99.9% = 0.1% allowed to fail; reframes 'never fail' into a number.
- [ ] **Page fast burn** — 1h burn ≥ 14.4× (budget gone in ~2 days) wakes the on-call now.
- [ ] **Ticket / review slow burn** — 6h burn ≥ 6× files a ticket; 3d burn ≥ 1× is a weekly review.

4 · SLAs & THE ON-CALL MODEL

- [ ] **Know the downtime** — 99.9% = ~43 min/mo; 99.99% = ~4.4 min/mo. Those minutes can land on premiere night.
- [ ] **Read the remedy** — a CDN credit (e.g. CloudFront 10%/25%) refunds your bill, not your lost revenue.
- [ ] **On-call rotation** — primary + secondary, short shifts; review pager load so the team does not burn out.
- [ ] **Severity + postmortem** — Sev-1/2/3 sets the response; blameless postmortems fix the system, not blame.

WHAT 99.9% ACTUALLY BUYS — AND WHY YOU RUN TIGHTER

A '99.9% uptime' SLA sounds airtight, but it permits real downtime: roughly 43 minutes a month, or about 8.8 hours a year — and nothing stops those minutes from landing during your biggest live event. Each extra nine cuts the allowance hard: 99.95% allows about 22 minutes a month, 99.99% about 4.4 minutes, 99.999% about 26 seconds — and each nine costs disproportionately more to build and operate, so most OTT services neither need nor can afford telco-grade availability. Just as important is what an SLA actually pays when it is missed. A typical CDN remedy is a service credit — Amazon CloudFront, for example, refunds 10% of the bill for monthly uptime between 99% and 99.9% and 25% below 99% — which returns a slice of what you paid the vendor, not the subscriber revenue, ad revenue, or goodwill you lost while viewers stared at a spinner. Treat an SLA as a billing-risk instrument, not insurance. The practical posture: pick the internal SLO the business genuinely needs, set the customer-facing SLA looser than that SLO so your own alarm trips before you breach a promise, run more than one CDN so a single vendor's bad night is a failover and not an outage, and read every vendor SLA for the exclusions and the credit math before you sign. Then back the numbers with people: a sustainable on-call rotation, clear severity levels, tested runbooks so the 3 a.m. response is procedure not improvisation, and blameless postmortems that turn each incident into a permanent fix instead of a name to blame.