

LL-HLS Readiness Checklist (2026)

Twenty-five items every team should verify before declaring an LL-HLS deployment production-ready.

Encoder & packager

- Segment duration set to 2–6 s; B-frames disabled or limited to one for low GOP latency.
- PART-TARGET set to 0.2 s (≤ 0.33 s for ultra-low-latency mode; ≤ 1.0 s for relaxed mode).
- At least one INDEPENDENT=YES part per second of real time (Apple authoring spec §2.16).
- Part durations sum exactly to #EXTINF when the segment is closed (verified with mediastreamvalidator).
- Packager produces CMAF chunks with one chunk per part — chunked-transfer-friendly.
- Encoder pipeline depth ≤ 1 s; production target ≤ 600 ms.

Playlist tags

- #EXT-X-VERSION ≥ 9 (required for #EXT-X-SKIP and full LL-HLS tag set).
- #EXT-X-SERVER-CONTROL present with CAN-BLOCK-RELOAD=YES, PART-HOLD-BACK $\geq 3 \times$ PART-TARGET, HOLD-BACK $\geq 3 \times$ target-duration.
- #EXT-X-PART-INF:PART-TARGET line at top of every LL-HLS media playlist.
- #EXT-X-PRELOAD-HINT line at bottom pointing at the next part URI.
- #EXT-X-RENDITION-REPORT line for every sibling video, audio, and subtitle rendition.
- CAN-SKIP-UNTIL $\geq 6 \times$ target-duration; #EXT-X-SKIP returned when _HLS_skip=YES.

Origin & long-poll behaviour

- Server actually long-polls _HLS_msn / _HLS_part requests (validated end-to-end, not just by playlist advertising).
- Response returns within ± 50 ms of the requested part being published.
- Discontinuities and #EXT-X-DATERANGE entries inside a skipped range are NEVER dropped from the response.
- CMAF init segments (#EXT-X-MAP) served separately and never affected by delta updates.

CDN & caching

- Cache key includes _HLS_msn, _HLS_part, _HLS_skip; playlist Cache-Control set to max-age=0 or 1.
- Parts cached aggressively (long TTL); playlists cached briefly; init segments cached for the whole stream.
- CDN supports HTTP/1.1 chunked transfer OR HTTP/2 stream framing OR HTTP/3 streams end-to-end.
- Origin shield in place to coalesce long-poll requests behind a single upstream connection.

Player & ABR

- ABR switch points constrained to INDEPENDENT=YES parts only.
- Rendition reports consumed to skip the probe round-trip on every ABR switch.
- Player playback buffer sized at 3–6 parts (not 3 segments); rebuffer policy tuned accordingly.
- Player long-polls the next part URI before parsing the previous response (request pipelining).

Validation & monitoring

- Apple mediastreamvalidator pass with -P flag against every variant; no warnings.
- Apple hlsreport generated for the master playlist; reviewed before every release.
- Per-region glass-to-glass latency monitored; alert above 6 s p95.