

MoQ Readiness Checklist

Before kicking off a Media over QUIC pilot — the architectural, vendor, and operational questions to answer first.

1. Product fit

- ☐ Does a sub-500 ms latency floor unlock a specific product feature (live betting market, live commerce buy moment, second-screen sync, interactive game stream)? If not — MoQ is the wrong call today.
- ☐ Are concurrent viewer counts in the thousands-to-millions range (where WebRTC SFU cost would dominate, or LL-HLS latency would be a product blocker)?
- ☐ Is the audience addressable through your own app or embed where you can ship a vendor MoQ player?
- ☐ Is the broadcast direction one-to-many or few-to-many (not bidirectional peer-to-peer like a video call)?
- ☐ Have you ruled out LL-HLS, LL-DASH/CMAF, and WebRTC delivery for the same workload with documented reasons?

2. Architecture decisions

- ☐ Encoder vendor: confirmed a 2026 MoQ origin (Bitmovin, Norsk, Wowza, Ant Media, Broadpeak, Synamedia, Red5, nanocosmos).
- ☐ Relay tier: decided between Cloudflare MoQ CDN (330+ cities, fastest pilot) and self-hosted moq-rs (full control, ops burden).
- ☐ Pinned to a specific draft revision (draft-ietf-moq-transport-17 or -latest); no "follow the head" in production.
- ☐ Streaming format: CMSF (CMAF over MoQ) chosen for the catalog/timeline; LOCMF alternative evaluated and noted.
- ☐ Player: vendor SDK identified; fallback path (LL-HLS or WebRTC) defined for non-supported devices (Apple Safari, smart TVs, older Android).
- ☐ GOP / object granularity: one CMAF chunk per object, NOT one segment per object — avoid the "HLS-like granularity" pitfall.

3. Operations & ops readiness

- ☐ UDP path tested across target regions; WebTransport fallback path tested for UDP-blocked corporate / mobile carrier networks.
- ☐ Capacity model: relay memory $\approx$ 200 bytes per active subscription (verify with your relay implementation); egress cost modelled at the CDN edge.
- ☐ Congestion behaviour: subgroup priority and partial-reliability semantics validated under packet loss (test 1%, 3%, 5%).
- ☐ Observability: per-track QoE metrics (start-up time, group join latency, stall rate) collected at the player and at the relay.
- ☐ DRM / encryption: in-transport QUIC encryption confirmed; end-to-end secure objects (draft-ietf-moq-secure-objects) needed if relay is untrusted.
- ☐ DoS posture: relay rate-limits per source IP; signed announcements; alignment with draft-englishm-moq-relay-dos guidance.

4. Vendor & spec risk

- ☐ Working Group Last Call timing tracked (RFC publication likely 6–12 months after WGLC); "upgrade plan" documented.
- ☐ Vendor SLA aligned to draft revision changes; spec breakage clause negotiated.
- ☐ Bridge to LL-HLS / WebRTC kept warm: if MoQ pilot fails, can the same audience be served by a fallback stack within 24 hours?
- ☐ Player auto-update path: forced refresh / soft-update infrastructure for catalog and message-format changes between draft revisions.
- ☐ Internal team: at least one engineer fluent in QUIC, WebTransport, MoQ data model; able to read the IETF draft directly.

Decision rule

If every box above is checked — proceed to a 6–8 week MoQ pilot against your defined product feature. If three or more boxes are unchecked in Sections 1–2 — stay on your current protocol and re-evaluate after the next moq-transport draft revision or after Working Group Last Call. If unchecked items are concentrated in Sections 3–4 — pilot on Cloudflare's MoQ CDN to limit your operational and spec-risk surface for the first iteration.