Learning-Analytics Pipeline Build Checklist

A one-page build checklist for the analytics pipeline: event schema, sampling, the LRS choice, the warehouse model, and the privacy controls to wire in from day one.

A. Capture - design the event schema first

- [ ] Choose the standard up front: xAPI (flexible) or Caliper (common in higher-ed)
- [ ] Define verbs and stable URL ids for every activity before building the player
- [ ] Adopt the xAPI Video Profile: capture played-segments and progress for video
- [ ] Set a registration per attempt; identify learners by an opaque account id, not email

B. Transport - batch, sample, survive the network

- [ ] Batch statements as an array; validate before flushing the atomic batch
- [ ] Sample: action verbs on user action plus a progress ping every ~10 seconds
- [ ] Never ship per-frame time-update capture to production
- [ ] Queue locally and flush on reconnect; statement-id makes retries idempotent

C. Store - choose or build the LRS

- [ ] Buy a conformant LRS unless scale or integration truly demands a build
- [ ] Confirm xAPI 2.0 (IEEE 9274.1.1-2023) conformance and the resources you need
- [ ] Treat the LRS as the immutable ledger, not the analytics database
- [ ] Weigh self-host (control) against commercial (managed scale and analytics)

D. Warehouse & privacy - model and protect

- [ ] Load raw first, then transform (ELT) into a star schema (fact + dimensions)
- [ ] Pre-aggregate heavy played-segments into per-cohort rollups
- [ ] Minimise data at capture; keep the identity map in a separate system
- [ ] Set retention and a delete-by-actor path before the first erasure request



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 checklist covers the event schema, batching and sampling, the LRS choice and the warehouse model. Our seven-volume handbook covers the video stack the events start in; start with Volume 3 for players and QoE data, then Volume 7 for data compliance and serving, and Volume 1 for the video ladder.

WHAT YOU WILL FIND INSIDE

- Player events and QoE monitoring (Vol. 3)
- Encoding ladders for course libraries (Vol. 1)
- Compliance: EU AI Act and biometric data (Vol. 7)

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

Building an xAPI pipeline from player to warehouse?

We will design the statement schema, pick or build the LRS and set up the ELT model with retention and erasure built in. 250+ video products shipped since 2005.

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