

# 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