

USE CASE 04 · FOR VP ENGINEERING · CTO · PLATFORM TEAMS

DROP-IN. MODERN RUNTIME. SAME API.

Your Elasticsearch clients, dashboards, ingest pipelines, and detection rules stay unchanged. What changes is the runtime beneath them — one Rust binary instead of a JVM cluster, 50 ms cold start instead of 15 seconds, 400 MB idle instead of 8.5 GB.

21× LESS MEMORY · 300× FASTER COLD START · SAME CLIENTS

ES RUNS THE WORLD. AT A COST.

Every enterprise already runs Elasticsearch. And every enterprise has the same three complaints: the license model changed, the JVM operations burden never shrinks, and the vector story arrived late and expensive. Meanwhile the AI transformation roadmap needs retrieval infrastructure that's faster, cheaper, and compatible with existing clients. Rewriting everything is not on the table.

LICENSING DRIFT	SSPL, feature gating, paid-only security tiers. The open-source ES era ended; the alternatives require migration. Your boards asked what the alternative looks like.
JVM OPERATIONS	Heap sizing. G1 vs ZGC tuning. Cold starts measured in seconds. Memory reserved up front and rarely returned. On-call runbooks half-filled with JVM incantations.
VECTOR BOLT-ON	Elasticsearch shipped kNN search, but vector workloads still compete with BM25 for heap. RAG teams move to Pinecone; the single-store advantage disappears.
MIGRATION RISK	Swapping a search engine is a year-long project. Every client library, every Logstash pipeline, every Kibana dashboard is a migration risk. Teams stay on the current ES because the alternative is worse.

03 · SAME API · DIFFERENT ENGINE

PORT :9200. YOUR CLIENTS. OUR ENGINE.

XERJ speaks the Elasticsearch 8.13 wire protocol. 5,165 LOC of compat layer covers 38 query types, every common aggregation, the bulk API, the scroll API, and the cluster-health endpoint. The client library you use today is the client library you use tomorrow. Migration is a configuration change, not a rewrite.

MEMORY

21×**LESS IDLE
RAM**

400 MB single node vs 8.5 GB for a 4-node ES cluster. RSS tracks working set — jemalloc returns pages.

COLD START

300×**FASTER
RESTART**

50 ms vs ~15 s JVM bootstrap. First query served when the socket binds — no classloader, no JIT warmup.

BINARY

56×**SMALLER
DEPLOY**

13 MB static binary vs ~620 MB JAR + JRE + config + policy files. Distroless base image.

COMPAT

ES 8.13**WIRE PROTOCOL
DROP-IN**

38 query types · bulk · scroll · cluster health · same response envelope. No client rewrite.

MIGRATION IS A CONFIG CHANGE · NOT A YEAR-LONG PROJECT

04 · DROP-IN BLUEPRINT

SAME CLIENTS. NEW ENGINE.

A three-node HA cluster that replaces a 4-node Elasticsearch cluster. Same client libraries continue pointing at :9200. Same dashboards, same detection rules, same Logstash pipelines.

CLIENTS · UNCHANGED

elasticsearch-py · elasticsearch-js · elasticsearch-java
Kibana · Grafana (ES datasource) · custom dashboards
Filebeat · Logstash · Vector · Fluentd
LangChain · LlamaIndex · Haystack – ES retriever

|

▼ HTTP :9200 · ES 8.13 wire protocol · same headers, same shapes

XERJ CLUSTER 3-node HA · same ingest / query API as ES

FTS index	BM25 · 38 query types · analyzers
VECTOR	HNSW · SQ8 · nibble · rerank
COLUMNS	doc-values · delta-ts · dictionary · varint
AGENT MEMORY	session-scoped · time-decay
WAL + mmap segment cache	lock-free readers

|

▼

MIGRATION MECHANICS

Dual-write	ingest pipeline writes to ES + XERJ in parallel
Shadow-query	application reads primary = ES · compare against XERJ
Cut-over	primary = XERJ · keep ES warm for N days
Decommission	ES cluster retired · contract terminated

SIZING

XERJ 1-node covers 4-node ES idle baseline.
XERJ 3-node HA matches 4-node ES on throughput + adds SLO headroom.

SAME :9200 · SAME DSL · SAME RESPONSE ENVELOPE

05 · WHERE THE DELTA COMES FROM

NO JVM. NO GC. NO DRIFT.

Five runtime properties that Elasticsearch cannot match without a rewrite it cannot do. Each is a measurable consequence of the Rust + tokio + mmap architecture, not a feature flag.

AOT STATIC BINARY	First byte served when the socket binds. No classloader scan, no JIT warmup, no policy-file verification. Cold start is 50 ms vs ~15 s — which matters for Kubernetes rolling restarts at fleet scale.
NO GC PAUSES	Ownership frees objects at scope exit. No tracing GC means no stop-the-world, no deopt cascade, no safepoint coordination. p99.9 tracks p50 plus cache-miss and disk — knowable, not tuned.
MMAP-SHARED CACHE	One OS page cache backs every reader. No repeated <code>mmap()</code> syscalls, no per-reader copy. A 30-day log retention stays hot without multiplying memory.
LOCK-FREE READERS	ArcSwap segment view + DashMap version map. Queries allocate nothing and acquire no mutex on the hot path. Flush swaps the view atomically; in-flight queries finish on the old view.
ONE PROCESS	Ingest parsers, query planner, and writer are the same tokio process. One <code>fsync</code> per batch, one network hop, end-to-end back-pressure. No Logstash box to babysit.
FORWARD- COMPAT SEGMENTS	Segment format versioned with a section table; older binaries skip unknown sections rather than fail. Rolling upgrades do not require full reindex.

06 · MEASURED VS ES 8.13

ONE NODE. FOUR-NODE PARITY.

Nine axes a procurement review looks at. XERJ on one node vs. a 4-node Elasticsearch 8.13 cluster doing the same SIEM + log workload.

AXIS	XERJ	ES 8.13 (4-NODE)
IDLE MEMORY	400 MB	8.5 GB — 21.3×
COLD START	50 ms	~15 s — 300×
RELEASE BINARY	13 MB	~620 MB — 56×
CONFIG KNOBS	47	3,000+ — 63.8×
BINARIES TO RUN	1	5 (ES + Logstash + Kibana +...)
DISK · 1 M SIEM EVENTS	85 MB	240 MB — 2.8×
SIEM TOP-SOURCE-IPS P95	0.4 ms	29.8 ms — 74×
NGINX TERMS AGG P95	0.4 ms	35.7 ms — 89×
CLIENT LIBRARY CHANGES	0	0 (ES native)
COST IMPACT	~80 % infrastructure cost reduction on log + search workloads, measured on a like-for-like corpus. Most of the saving is headcount: one team owns the data layer, not four.	
SOURCE	SIEM_BATTLE_2026-04-14_184900.UTC.md · CLUSTER_BATTLE_2026-04-14_180241.UTC.md · FAIR_BATTLE_V3_20GB_2026-04-14.md — reports ship with reproduction scripts.	

07 · EVERYTHING ES · UNCHANGED

SAME CLIENTS.
SAME INGEST.
SAME DASHBOARDS.

The promise is that nothing in your stack above port 9200 changes. Here's the list.

CATEGORY	TOOLS	STATUS
CLIENT LIBRARIES	elasticsearch-py · elasticsearch-js · elasticsearch-java · Go · .NET · Ruby · PHP · Rust — all via ES wire	GA
INGEST	Filebeat · Winlogbeat · Auditbeat · Metricbeat · Packetbeat · Logstash · Vector · Fluentd · Fluent Bit	GA
DASHBOARDS	Kibana (community-validated) · Grafana with ES datasource · any custom dashboard using ES HTTP API	GA
SEARCH UIS	Search UI frameworks built on ES · Relevance Workbench equivalents · App Search replacements	GA
RAG / AI	LangChain ES retriever · LlamaIndex ES backend · Haystack · DSPy — zero adapter code	GA
SIEM TOOLS	Wazuh · Elastic Security · Sigma rule converters · detection-as-code frameworks	GA
SCROLL / PIT	Point-in-time snapshots + scroll cursors · ES-compatible semantics · same response envelope	GA
IDENTITY	OIDC · SAML 2.0 · AWS IAM · Azure AD · Okta · SCIM provisioning · per-tenant indices	ROAD Q2
X-PACK FEATURES	RBAC · audit log · ILM · ML anomaly detection — covered by XERJ-native equivalents or on roadmap	ROAD Q2-Q4

08 · HOW MIGRATION STAYS SAFE

DUAL-WRITE. SHADOW-READ. REVERSIBLE.

ES-wire compatibility means the migration is reversible at every step. If the XERJ pilot fails on a benchmark, rolling back is a config change — the original ES cluster stays primary until the team commits.

DUAL-WRITE	Ingest pipeline writes to ES primary + XERJ shadow in parallel. Same batch, same _bulk call, two destinations. No code change; one Logstash/Vector config line.
SHADOW-READ	Application reads from ES primary. A sampled fraction of queries is replayed against XERJ for latency, correctness, and recall comparison. No user impact.
A/B SWITCH	One env var flip routes N % of read traffic to XERJ. Start at 1 %, ramp as confidence grows. Monitoring shows p95, error rate, and recall drift side by side.
CUT-OVER	Primary becomes XERJ. ES stays warm as rollback target for N days. Once decommissioned, the ES license terminates and the 4-node JVM cluster is reclaimed.
ROLLBACK	One env var flip routes back to ES. Segments created on XERJ during the primary window are snapshot-exportable to ES-readable format (roadmap Q3). Worst case: re-ingest from the hot source.
DATA MIGRATION	Historical data: snapshot + re-ingest is fastest at most scales. Dedicated migration tooling lands Q2 (ES snapshot → XERJ segment). Re-embedding required for vector workloads.

NO LOCK-IN · ES WIRE COMPATIBILITY GOES BOTH WAYS

09 · FIVE PHASES · ONE YEAR

MIGRATE WITHOUT RISK.

A typical enterprise ES-replacement rollout. Low-risk, reversible, and measured at every stage.

01	PILOT WEEKS 1-2	Reproduce the SIEM_BATTLE and CLUSTER_BATTLE reports on your own corpus. Private binary, seeded data, reproduction scripts. Your numbers, your hardware, your decision.
02	SHADOW WEEKS 3-6	Run XERJ alongside ES. Dual-write · sampled shadow-read · Kibana on both. Compare latency, memory, correctness on production traffic. Still no user impact.
03	CUT-OVER ONE WORKLOAD QUARTER 2	Pick the smallest workload — usually observability logs or SIEM. Primary flips to XERJ. ES warm-standby for rollback. First bill line removed.
04	ADD VECTOR + MEMORY QUARTER 3	Move the vector DB and session cache onto the same platform. Hybrid retrieval lights up. RAG orchestrators simplify. One team owns the retrieval layer now.
05	DECOMMISSION ES QUARTER 4	Final workloads migrated. ES cluster retired. Contract terminated. One platform, one vendor, one SLA, one audit surface.

REVERSIBLE AT EVERY PHASE · ES WIRE COMPATIBILITY MEANS NO LOCK-IN

10 · NEXT STEPS

RUN IT. ALONGSIDE ES.

A two-week shadow pilot. Private binary. Repro scripts for SIEM_BATTLE and CLUSTER_BATTLE. Config templates for Filebeat / Vector dual-write.

PLATFORM CONTACT

HELLO@XERJ.AI

XERJ.AI · XERJ.COM · XERJ.DEV

WHAT YOU GET	Private binary · SIEM_BATTLE + CLUSTER_BATTLE repro · Filebeat / Vector dual-write config · Kibana dashboard JSON · migration runbook.
WHAT YOU BRING	One existing ES-backed workload · its current client library list · a latency baseline for at least one query you care about.
COMPANION BRIEFS	xerj-tech-brief.pdf · xerj-exec-brief.pdf · xerj-usecase-security.pdf · xerj-usecase-rag.pdf