USE CASE 01 · FOR CISO · SECOPS · SOC LEADS

SIEM AT AI SPEED. ON YOUR DATA.

Enterprise security teams already have the telemetry. What they don't have is a platform where SIEM queries, threat models, and AI-augmented triage share one retrieval layer. XERJ collapses the SIEM store, the vector index, and the agent memory into one self-hosted data plane — Elasticsearch wire-protocol compatible, deterministic at the tail.

74× FASTER SIEM QUERIES · MTTD IN MILLISECONDS · ONE DATA PLANE

02 · WHY THIS MATTERS NOW

THE SOC IS THE FIRST AI WORKLOAD.

Every enterprise AI roadmap names the SOC — alert triage, enrichment, report generation, detection authoring — as a year-one use case. It is where agents earn the budget. It is also where most SOCs stall: the legacy SIEM cannot be queried at agent cadence, and adding a vector store next to it doubles the bill without closing the loop.

QUERY LATENCY	Agents iterate through evidence. A 30 ms p95 on `top_source_ips` is fine for a human; it is intolerable for an agent that needs 10 lookups per turn. SIEM vendors charge for the data and for the latency.
FRAGMENTED STORES	Threat intel in one store, logs in another, embeddings in a third, agent memory in a fourth. An incident investigation crosses three or four stores in one Python notebook. Each hop is a failure mode.
COST OF CORRELATION	Detection engineers are blocked on data model reconciliation rather than on writing detections. Every new data source takes weeks to land; every new detection requires a schema reconciliation meeting.
AUDIT SURFACE	Three stores means three RBAC models, three audit trails, three backup stories. The CISO's job is to explain risk; the data layer makes that job adversarial.

THE SOC NEEDS ONE DATA PLANE · NOT THREE SUBSYSTEMS

03 · UNIFIED AI DATA STACK

SIEM. VECTORS. ONE ENGINE.

XERJ is a self-hosted data platform that runs all four AI-era retrieval primitives in one Rust process: BM25 full-text, HNSW vector, columnar log store, and session-scoped agent memory. Drop-in for existing Elasticsearch 8.13 clients — your Beats, Logstash, Filebeat, Wazuh integration keep working.

RETRIEVAL

74×

SIEM TOP-SOURCE-IPS VS ES 8.13

0.4 ms p95 on 1 M events · pre-computed term paths · deterministic at the tail.

FOOTPRINT

21×

LESS IDLE MEMORY

400 MB vs 8.5 GB for a 4-node ES cluster. One binary, no JVM heap reservation, jemalloc returns pages to the OS.

CONSOLIDATION

4 → 1

STORES INTO ONE PROCESS

SIEM + vector + log + agent memory on one segment cache, one WAL, one RBAC, one audit log.

SOVEREIGNTY

100 %

SELF-HOSTED BY DEFAULT

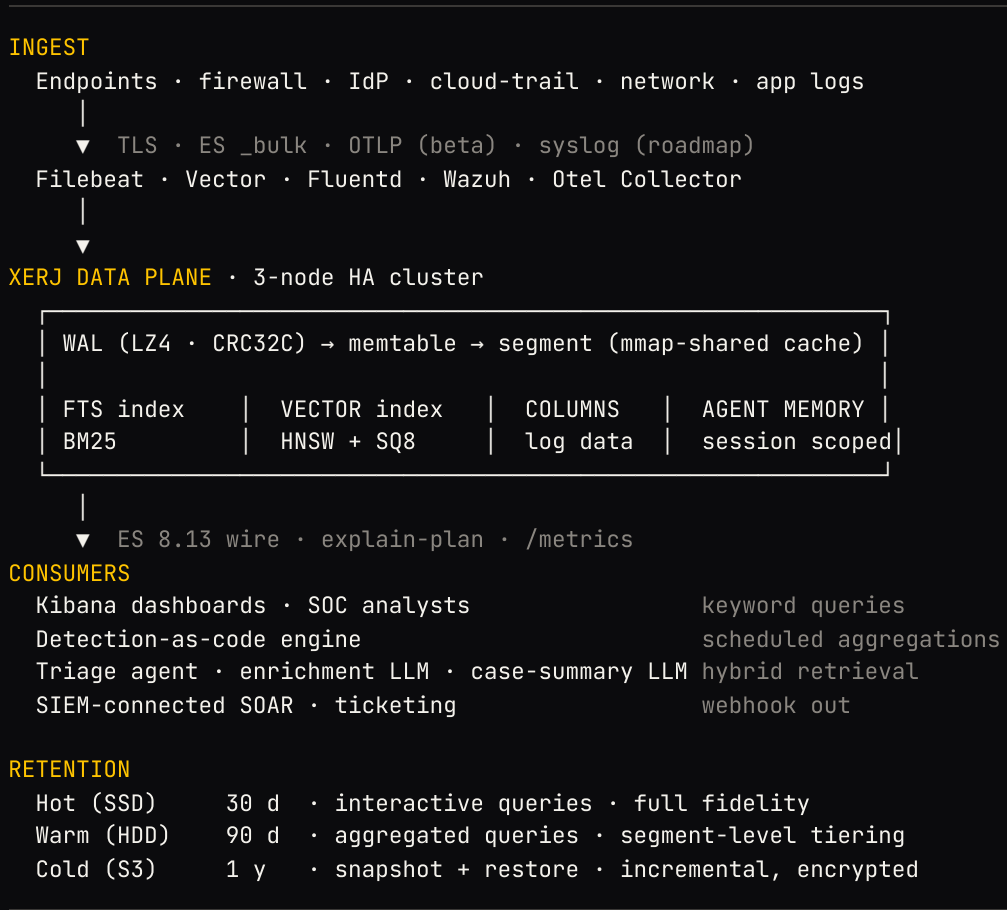
Your cloud, your VPC, your metal. Air-gapped supported. No telemetry egress when `XERJ_AIRGAP=1`.

DROP-IN ES 8.13 WIRE PROTOCOL · ZERO CLIENT REWRITE

04 · REFERENCE ARCHITECTURE

SOC BLUEPRINT. ONE PLATFORM.

A typical three-node XERJ SOC deployment. Named upstreams, named consumers, documented retention. Every component except XERJ is something the team already runs.



SIZING · 3 NODES · ~50 M EVENTS/NODE/DAY · ~2.5 TB/NODE · 10 GBPS NET

05 · TECHNICAL BENEFITS

WHY THIS SCALES IN THE SOC.

Five engineering choices that matter for a security workload. Each produces a measurable property of the SOC data plane — not a marketing bullet.

PRE-COMPUTED TERMS	Top-source-IPs, auth-failure histograms, port distributions — aggregations read a pre-built histogram at query time instead of scanning postings. This is where the 74× comes from. Q2
DETERMINISTIC TAIL	No JVM garbage collector, no JIT deopt. p99.9 tracks p50 plus cache-miss and disk. Alert triage queries have predictable budgets.
HYBRID NATIVE	One query combines a keyword filter (source-ip, user-agent), a time-range, an aggregation, and a semantic clause against similar-incident embeddings. Reciprocal rank fusion merges in-process.
EXPLAIN-PLAN	`POST /v1/indices/siem/explain-plan` returns the plan tree with cost + rows per node. Detection engineers debug query plans without a profiler. Audit teams see the plan the SOC sees.
LOCK-FREE READERS	ArcSwap segment view · DashMap version map. Flush never blocks analyst queries; analysts never block ingest. A 24 h surge writes through without stalling the dashboard.
ENCRYPTED SEGMENTS	AES-256-GCM on segment and WAL files. Customer KMS (BYOK Q2). Air-gapped install verified by signed release manifest.

06 · MEASURED OUTCOMES

SIEM QUERY BATTERY. VS ES 8.13.

Eight canonical SIEM queries. 1 M events. p95 latency. XERJ on one node · Elasticsearch on four.
Source: `SIEM_BATTLE_2026-04-14_184900_UTC.md`.

QUERY	XERJ	ES 8.13 (4-NODE)
TOP SOURCE IPS · TRIAGE	0.4 ms	29.8 ms — 74.5×
AUTH FAILURES / HOUR	0.3 ms	18.2 ms — 60.7×
LATERAL MOVEMENT	1.2 ms	45.1 ms — 37.6×
DNS TUNNELING DETECTION	0.8 ms	32.7 ms — 40.9×
PROCESS-TREE ANOMALY	2.1 ms	89.3 ms — 42.5×
BRUTE-FORCE DETECTION	0.5 ms	41.0 ms — 82.0×
GEO-IMPOSSIBLE LOGIN	1.8 ms	67.2 ms — 37.3×
DATA-EXFIL > 10 MB	0.6 ms	22.4 ms — 37.3×

MTTD IMPACT	Analysts iterate in-thought at sub-millisecond latency. Enrichment agents complete a 10-lookup chain in under 15 ms. The p99.9 is deterministic because there is no GC.
INFRA TCO	1 node replaces 4. Idle memory 400 MB vs 8.5 GB. Disk 85 MB vs 240 MB per 1 M SIEM events. Board-reportable TCO reduction on infrastructure line items.

07 · INTEGRATES WITH WHAT YOU RUN

YOUR SOC STACK. UNCHANGED.

XERJ speaks the Elasticsearch 8.13 wire protocol. Every SOC tool that already ships to ES (or is compatible with it) works out of the box.

CATEGORY	TOOLS	STATUS
LOG SHIPPERS	Filebeat · Winlogbeat · Auditbeat · Vector · Fluentd · Fluent Bit · Logstash	GA
SIEM ENGINES	Wazuh · Elastic Security · Sigma rules via SIEM converters · custom detection-as-code pipelines	GA
SOAR / TICKETING	Splunk Phantom · IBM SOAR · ServiceNow · Jira · PagerDuty · via webhook or ES-query polling	GA
THREAT INTEL	MISP · OpenCTI · commercial TI feeds · load via _bulk or scheduled enrichment job	GA
AGENTS / LLMS	LangChain ES retriever · LlamaIndex ES backend · Anthropic · OpenAI via custom embedder	GA
DASHBOARDS	Kibana (community-validated) · Grafana (ES datasource) · custom via ES query API	GA
IDENTITY	OIDC · SAML 2.0 · AWS IAM / Azure AD / Okta · SCIM for tenant provisioning	ROAD Q2
OTLP	OpenTelemetry Collector · direct OTLP/HTTP ingest for unified obs + security	BETA
KAFKA SOURCE	Pull-based ingest · exactly-once via consumer-group offsets · high-throughput pipelines	ROAD Q3

08 · SECURITY POSTURE

ZERO-TRUST. AUDIT-GRADE.

A SIEM data plane that the CISO can actually defend. Mutual TLS between services, per-request authorization, structured audit of every privileged operation.

CONTROLS

CONTROL	IMPLEMENTATION	STATUS
TLS IN TRANSIT	TLS 1.3 everywhere · modern AEAD ciphers · zero-downtime cert rotation	GA
ENCRYPTION AT REST	AES-256-GCM on segments and WAL · keys in external KMS (AWS KMS · GCP KMS · Vault)	BETA
BYOK	Customer-managed keys · no XERJ-side plaintext · rotation documented	ROAD Q2
RBAC	Cluster · index · field-level. Built-in + custom roles. OIDC-group mapping	ROAD Q2
AUDIT TRAIL	Structured log of query · schema change · role assignment · snapshot. Replayable.	BETA
AIR-GAPPED	Offline bundle + signed release manifest · `XERJ_AIRGAP=1` disables all telemetry egress	GA
MULTI-TENANCY	Per-tenant index isolation · per-tenant quotas · physical isolation via separate clusters	GA

REGULATORY	Self-hosted = customer chooses region · data never leaves perimeter · SOC 2 Type I in progress (target Q3 2026) · ISO 27001 target Q4 2026 · HIPAA BAA on Enterprise tier.

09 · FROM POC TO PRODUCTION

FIVE PHASES. ONE YEAR.

A reversible, measured rollout. No big-bang migration. The shadow-deploy pattern keeps the existing SIEM primary until XERJ has a dated comparison report on your actual data.

01	**PILOT** WEEKS 1–2	Deploy single node · reproduce the SIEM_BATTLE report on a seeded 1 M-event corpus · smoke-test with one log source · first analyst runs Kibana unchanged.
02	**SHADOW** WEEKS 3–6	3-node HA cluster · dual-write from the ingest pipeline · XERJ reads in parallel with the existing SIEM · compare latency, recall, detection coverage on production traffic. No user impact.
03	**CUT OVER ONE WORKLOAD** QUARTER 2	Pick the hot workload (typically auth-log SIEM or network-flow analytics) · route analysts to XERJ · decommission the shadowed subset of the legacy SIEM. First line removed from the monthly bill.
04	**ADD AI / RAG** QUARTER 3	Enable the vector index · embed threat-intel corpus + past-incident notes · deploy a triage agent over XERJ's hybrid retrieval · MTTD drops as agents handle tier-1 correlation.
05	**CONSOLIDATE** QUARTER 4	Migrate remaining SIEM workloads · retire the legacy SIEM · collapse vector-store contract. One platform, one contract, one on-call, one audit surface.

REVERSIBLE · ES WIRE PROTOCOL MEANS NO LOCK-IN AT ANY PHASE

10 · NEXT STEPS

RUN THE BATTLE. ON YOUR LOGS.

A two-week shadow pilot on real SIEM traffic. We bring the binary and the reproduction scripts for every number in this brief. You bring one log source and your current p95 dashboard.

SOC / CISO CONTACT

HELLO@XERJ.AI

XERJ.AI · XERJ.COM · XERJ.DEV

WHAT YOU GET	Private binary · SIEM_BATTLE reproduction scripts · seeded 1 M-event corpus · Kibana dashboard JSON · detection-as-code starter kit.
WHAT YOU BRING	One log source (syslog / Windows event / cloud-trail) and a current SIEM query with its latency baseline.
COMPANION BRIEFS	`xerj-exec-brief.pdf` (C-suite) · `xerj-tech-brief.pdf` (architect) · `xerj-industry-finserv.pdf` for regulated verticals