

FOR ENTERPRISE ARCHITECTS · PLATFORM LEADS · PRINCIPAL ENGINEERS

# THE UNIFIED DATA PLANE FOR THE AGENT ERA.

XERJ is a self-hosted data platform that consolidates full-text search, vector retrieval, columnar logs, and agent memory into a single Elasticsearch-compatible service. Drop-in for existing ES clients; purpose-built for enterprises standing up retrieval-augmented applications at scale.

ES 8.13 COMPATIBLE · SELF-HOSTED · DATA-SOVEREIGN · ZERO-TRUST READY

## 02 · EXECUTIVE SUMMARY

# ONE PLATFORM. EVERY RETRIEVAL WORKLOAD.

Enterprise AI programs stall when every team assembles a different retrieval stack: ES for logs, Pinecone for vectors, Redis for session memory, a stitching orchestrator for every hybrid query. XERJ collapses that stack into one platform under one API. Migration is a configuration change — the application keeps its Elasticsearch clients.

## UNIFIED DATA PLANE

## FOUR WORKLOADS. ONE ENGINE.

Full-text (BM25), dense-vector (HNSW), columnar logs, and agent memory in a single process, single segment cache, single WAL. No cross-store orchestrator.

## DROP-IN ES COMPATIBILITY

## ZERO REWRITE. PORT :9200.

ES 8.13 wire protocol · 38 query types · bulk + scroll + search APIs. Existing clients, dashboards, and ingest pipelines work unchanged.

## ENTERPRISE OPERATIONS

## SOVEREIGN. AUDIT-READY.

Self-hosted on your cloud or metal. Per-query explain-plan with cost + row counts. Prometheus metrics first-class. RBAC roadmap [Q2](#).

## MODERN RUNTIME

## RUST. NO JVM.

13 MB static binary · 50 ms cold start · 400 MB idle · no GC pauses. Deterministic tail latency — SLOs mean something in the 99.9 percentile.

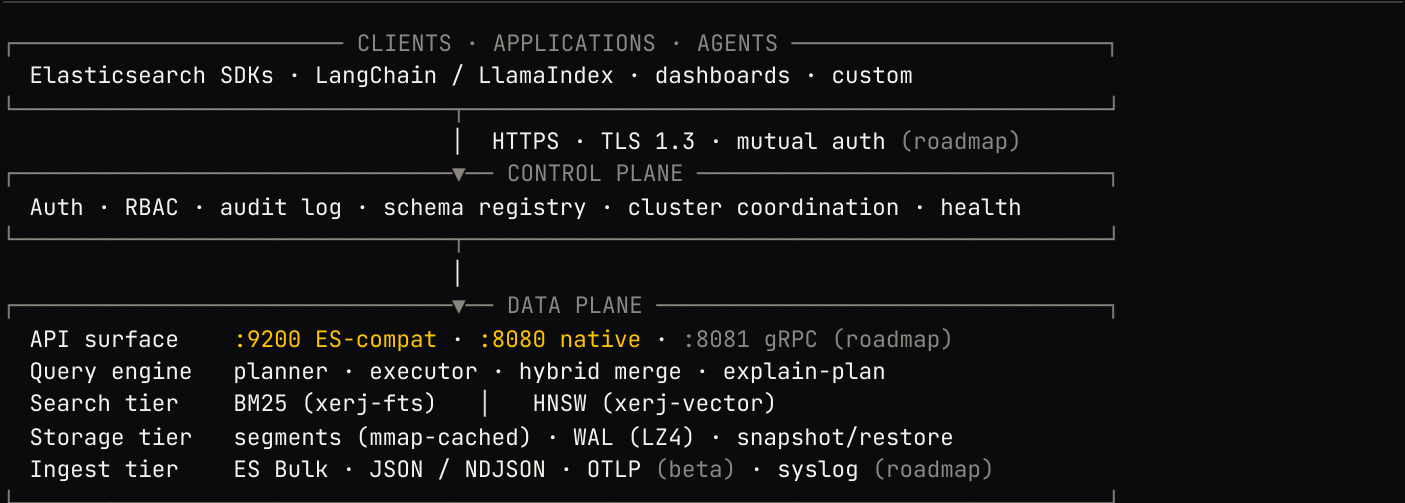
## PRIMARY USE CASES

SIEM MODERNIZATION · RAG AT ENTERPRISE SCALE · UNIFIED OBSERVABILITY · AGENT MEMORY PLATFORM · ELASTICSEARCH REPLACEMENT

03 · SOLUTION ARCHITECTURE

# CONTROL PLANE. DATA PLANE.

XERJ separates control-plane (cluster state, schema, RBAC, audit) from data-plane (ingest, storage, query). The data plane scales horizontally; the control plane stays small and highly available. Both are observable through a single Prometheus endpoint.



CROSS-CUTTING CONCERNS

|               |                                                              |
|---------------|--------------------------------------------------------------|
| Security      | TLS · RBAC (roadmap) · BYOK (roadmap) · tenant isolation     |
| Observability | Prometheus · OTLP out (roadmap) · structured JSON logs       |
| Reliability   | WAL + checkpoint · segment snapshots · replication (cluster) |
| Governance    | per-query explain-plan · per-index retention · audit trail   |

KEY PROPERTY · DATA PLANE AND CONTROL PLANE SCALE INDEPENDENTLY · NO SHARED HOT PATH

## 04 · DEPLOYMENT TOPOLOGIES

# FOUR PATTERNS. FOUR ENVIRONMENTS.

Choose the topology that fits the workload's SLO, retention horizon, and regulatory scope. All four patterns use the same binary and the same configuration surface — the difference is how many nodes and where they live.

## SINGLE-NODE

## POC · DEV

One binary, one host. Full feature set. Ideal for pilots, CI pipelines, edge sites, and air-gapped labs.

1 node  
RT0: N/A (single point)  
RP0: WAL fsync  
Capacity: up to ~250 M docs

## HA CLUSTER

## STANDARD · PROD

Three nodes, one region, synchronous replication. Typical enterprise production footprint.

3 nodes · 1 region  
RT0: < 30 s  
RP0: 0 (sync)  
Capacity: 1-5 B docs

MULTI-REGION ROAD Q3

## GLOBAL · RESILIENT

Active-active across regions. Async cross-region replication. Failover via control-plane DNS swap.

3 + 3 nodes · 2 regions  
RT0: < 2 min  
RP0: async (sub-sec)  
Capacity: > 10 B docs

## AIR-GAPPED

## ON-PREM · REGULATED

No internet. Offline artifact bundle · license file · offline update channel. Deployable to a sealed rack or SCIF.

1-N nodes  
No telemetry egress  
Offline signed updates  
Target: GOV · HEALTH · FIN

## BYOC KUBERNETES

## CLOUD · NATIVE

Helm chart + CRDs. StatefulSet per data-plane role. PVCs for segments; PV for WAL. Tested on EKS, GKE, AKS.

Helm · Terraform ROAD Q2  
StorageClass: fast-SSD  
HPA on CPU + ingest rate  
Pod-level audit

MANAGED SAAS ROAD '27

## HOSTED · TURNKEY

Multi-tenant control plane, single-tenant data plane. Predictable pricing per GB ingested + query volume.

Dedicated data plane  
Shared control plane  
Region-of-choice  
BYOK supported

05 · UNIFIED DATA PLATFORM

# FOUR STORES. ONE ENGINE.

Every store type shares a single segment cache, a single WAL, a single permission model, and a single query language. The distinctions below are logical — not separate deployments, not separate bills, not separate on-call rotations.

| STORE TYPE    | CAPABILITIES                                                                                                                            | STATUS  |
|---------------|-----------------------------------------------------------------------------------------------------------------------------------------|---------|
| FULL-TEXT     | BM25 scoring · 38 ES query types · 5+ analyzers · per-field overrides · position-aware postings for phrase queries                      | GA      |
| VECTOR        | HNSW k-NN · cosine / L2 / inner-product · SQ8 (4×) & nibble (8×) quantization · optional float32 re-rank · 4,096 dims                   | GA      |
| COLUMNAR LOGS | Columnar doc-values · delta-of-delta timestamps · dictionary + ZSTD keywords · time-partitioned memtables · terms / date-histogram aggs | GA      |
| AGENT MEMORY  | Session-scoped indices · time-decay scoring · token-budget-aware retrieval · native hybrid with vector + keyword signals                | BETA    |
| METRICS       | Prometheus remote_write ingest · downsampling · retention tiers · range queries over columnar store                                     | ROAD Q3 |
| TRACES        | OTLP ingest · span joins · service-graph materialization · trace-to-log pivot via shared schema                                         | ROAD Q4 |

## WHY UNIFIED

|                     |                                                                                                                                                               |
|---------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|
| DATA GRAVITY        | Logs, embeddings, and metrics describe the same events. Keeping them in one place means queries join at the storage layer, not through a Python orchestrator. |
| OPERATIONAL SURFACE | One binary to deploy. One config to tune. One audit trail. One backup story. One on-call rotation.                                                            |
| COST                | Four vendors collapse to one. Storage is shared across store types — a log index and a vector index live in the same segment file when schemas allow.         |

06 · DATA INGESTION

# MANY SOURCES. ONE PIPELINE.

Every ingest path terminates at the same WAL and the same memtable. Durability is guaranteed at fsync return; visibility follows on flush. Back-pressure propagates to the producer so clients never silently drop.

## SUPPORTED PROTOCOLS & FORMATS

| PROTOCOL / FORMAT       | USE CASE                                                                                        | STATUS  |
|-------------------------|-------------------------------------------------------------------------------------------------|---------|
| ES _BULK API            | Drop-in for existing Beats, Logstash, Filebeat pipelines. Newline-delimited action + doc pairs. | GA      |
| JSON / NDJSON           | Direct HTTP ingest for app logs, custom producers, agent traces. One object per line.           | GA      |
| OTLP / HTTP             | OpenTelemetry logs + traces. Standard collectors (Otel Collector, Vector) without a translator. | BETA    |
| SYSLOG (RFC 5424)       | Legacy infrastructure ingest. UDP + TCP listeners, per-source parsing rules.                    | ROAD Q2 |
| PROMETHEUS REMOTE_WRITE | Metrics ingest compatible with existing scrapers. Maps series to columnar store.                | ROAD Q3 |
| KAFKA SOURCE            | Pull-based ingest from Kafka topics. Exactly-once via consumer-group offset commits.            | ROAD Q3 |
| S3 / GCS / BLOB         | Batch import from object storage. Daily log archives, parquet snapshots, embedding dumps.       | ROAD Q4 |

## RELIABILITY GUARANTEES

|                 |                                                                                                                                                     |
|-----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------|
| DURABILITY      | Every accepted document is persisted to WAL with fsync + CRC32C before ACK. Crash recovery replays from the last checkpoint, idempotent per seq_no. |
| BACK-PRESSURE   | Ingest returns 429 / Retry-After when the WAL queue exceeds the watermark. Producers slow down; no silent drops.                                    |
| SCHEMA HANDLING | Segment-scoped schema. Field additions take effect on the next memtable flush — no global reindex, no service restart.                              |

## 07 · SEARCH &amp; RETRIEVAL

# KEYWORD. VECTOR. ONE QUERY.

Hybrid retrieval is the default, not a feature. A single query can express a keyword filter, a semantic clause, an aggregation, and a time-window constraint — the planner composes and executes them in one pass.

| CAPABILITY      | DETAIL                                                                                                                      | STATUS  |
|-----------------|-----------------------------------------------------------------------------------------------------------------------------|---------|
| FULL-TEXT       | BM25 · k1 / b tunable · 38 ES query types (term · match · bool · phrase · fuzzy · wildcard · prefix · multi_match · ...)    | GA      |
| VECTOR SEARCH   | HNSW k-NN · per-index quantization choice · runtime-tunable ef_search trading recall for latency                            | GA      |
| HYBRID MERGE    | Reciprocal Rank Fusion (RRF) — ranks from keyword + vector combined without magnitude calibration. In-process, single pass. | GA      |
| UNIFIED PLANNER | Shared cost model across FTS, vector, and aggregations. Joint top-K cut-off across signals.                                 | ROAD Q3 |
| AGGREGATIONS    | terms · date_histogram · histogram · percentiles · composite · stats · avg / sum / min / max. ES-compatible shapes.         | GA      |
| EXPLAIN PLAN    | Every query plan is queryable via POST /v1/indices/{i}/explain-plan. Per-node cost + row counts. First-class endpoint.      | GA      |
| SCROLL / PIT    | Point-in-time snapshots + scroll cursors for large result sets. ES-compatible semantics.                                    | GA      |
| SQL OVER HTTP   | ANSI-SQL frontend over the same storage. BI tools connect via standard JDBC / ODBC.                                         | ROAD Q4 |

## GOVERNANCE & COST CONTROL

|                  |                                                                                                                                                                  |
|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| PER-QUERY COST   | The explain-plan returns CPU, I/O, and row-scan estimates before execution — and realized counts after. Quotas and circuit-breakers enforced at the API.         |
| CIRCUIT-BREAKERS | Per-request heap · per-request time · per-tenant concurrency · per-index doc-scan ceiling. Triggers return a structured error with the budget that was exceeded. |

## 08 · INTEGRATION ECOSYSTEM

# YOUR STACK. UNCHANGED.

XERJ's integration strategy is to implement standard protocols rather than ship bespoke adapters. Anything that speaks Elasticsearch 8.13 or OpenTelemetry works out of the box.

## CLIENT LIBRARIES

elasticsearch-py · elasticsearch-js ·  
elasticsearch-java · via ES wire  
Go, .NET, Rust – any OSS client targeting ES 8.x

## DASHBOARDS & BI

Kibana · ES backend, community-validated  
Grafana · ES datasource · Prometheus scrape  
Tableau · Power BI · SQL frontend (roadmap Q4)

## IDENTITY & ACCESS

OIDC / SAML 2.0 · roadmap Q2  
SCIM provisioning · roadmap Q3  
AWS IAM / Azure AD / Okta · roadmap Q2

## DEPLOYMENT & AUTOMATION

Helm chart · Kubernetes Operator · roadmap Q2  
Terraform module · roadmap Q2  
Ansible playbooks · community

## INGEST COLLECTORS

Beats · Filebeat · Logstash · Vector · ES output  
OpenTelemetry Collector · OTLP (beta)

## AI & RAG FRAMEWORKS

LangChain · LlamaIndex · ES retriever  
Haystack · DSPy · ES backend  
Custom embedders via pluggable provider interface

## OBSERVABILITY

Prometheus · /metrics first-class  
OTLP export · roadmap Q3  
Structured JSON logs · stdout by default

## BACKUP & DR

S3 / GCS / Azure Blob · snapshot target  
Restic / Kopia · community, file-level  
Cross-region replication · roadmap Q3



## 09 · SECURITY &amp; COMPLIANCE

# ZERO-TRUST READY.

## AUDIT-GRADE.

The security model assumes a hostile network. Mutual TLS between services, per-request authorization, and a structured audit log of every privileged operation. Compliance certifications are on the roadmap; the engineering controls that those certifications require are either shipped or in flight.

### CONTROLS

| CONTROL            | IMPLEMENTATION                                                                                                         | STATUS     |
|--------------------|------------------------------------------------------------------------------------------------------------------------|------------|
| TLS IN TRANSIT     | TLS 1.3 on all listeners. Cipher suites restricted to modern AEAD. Certificate rotation without restart.               | GA         |
| ENCRYPTION AT REST | AES-256-GCM on segment and WAL files. Keys in an external KMS (AWS KMS, GCP KMS, HashiCorp Vault).                     | BETA       |
| BYOK               | Customer-managed keys via KMS. No XERJ-side plaintext access. Rotation documented.                                     | ROAD<br>Q2 |
| RBAC               | Role-based access at cluster, index, and field level. Built-in roles + custom roles. Integrated with OIDC groups.      | ROAD<br>Q2 |
| AUDIT TRAIL        | Structured log of every privileged operation: query, schema change, role assignment, snapshot. Replayable, exportable. | BETA       |
| MULTI-TENANCY      | Tenant-isolated indices, metrics, and quota. Per-tenant RBAC. Physical isolation via separate clusters.                | GA         |
| AIR-GAPPED UPDATES | Offline bundle + signed release manifest. Verified on install. No telemetry egress when XERJ_AIRGAP=1.                 | GA         |

### CERTIFICATIONS & ATTESTATIONS

| STANDARD              | SCOPE                                                                                | STATUS              |
|-----------------------|--------------------------------------------------------------------------------------|---------------------|
| SOC 2 TYPE I          | Platform controls · access management · change management · incident response        | ROAD<br>Q3<br>2026  |
| SOC 2 TYPE II         | 12-month operating effectiveness. Follows Type I.                                    | ROAD<br>Q2<br>2027  |
| ISO 27001             | ISMS scope covers XERJ platform engineering + SaaS operations                        | ROAD<br>Q4<br>2026  |
| HIPAA / HITECH        | BAA available on Enterprise tier. Self-hosted deployments inherit customer controls. | ROAD<br>2027        |
| GDPR / DATA RESIDENCY | Self-hosted = customer chooses region. SaaS: region-pinned data plane.               | GA<br>(SELF-HOSTED) |

10 · SCALABILITY & PERFORMANCE

MEASURED.  
REPRODUCIBLE.  
SCALE-OUT.

Every number below comes from a dated, checked-in battle report under engine/releases/v0.1.0/reports/ with a reproduction script. Scale rows above 1 node are projections pending the Q3 sharded-ingest GA.

PERFORMANCE PROFILE · PER NODE

| METRIC                      | VALUE      | AT SCALE       | VS ES 8.13 | SOURCE         |
|-----------------------------|------------|----------------|------------|----------------|
| Cold start                  | 50 ms      | same           | 300×       | SIEM_BATTLE    |
| Idle RSS                    | 400 MB     | +~200 MB / idx | 21×        | SIEM_BATTLE    |
| Binary size                 | 13 MB      | static         | 56×        | release        |
| SIEM top-source-IPs         | 0.4 ms p95 | 1 M events     | 74×        | SIEM_BATTLE    |
| NGINX terms agg             | 0.4 ms p95 | 5 distinct     | 89×        | CLUSTER_BATTLE |
| Hybrid query p95            | 38 ms      | 1 B vectors    | n/a        | hybrid_bench   |
| Disk · 1 M SIEM events      | 85 MB      | post-ingest    | 2.8×       | SIEM_BATTLE    |
| Vector memory · 10 M × 1536 | 18 GB      | SQ8            | 5.1×       | RAG_BENCH      |

CAPACITY-PLANNING HEURISTIC

|               |                                                                                                                  |
|---------------|------------------------------------------------------------------------------------------------------------------|
| SIEM WORKLOAD | ~50 M events / node / day sustained · hot retention 30 d · 1 node covers ~1.5 B events / 85 GB disk.             |
| RAG WORKLOAD  | ~55 M × 768-dim embeddings / node at SQ8 · query p95 < 40 ms · rebuild HNSW at 10 % churn.                       |
| OBSERVABILITY | ~80 K events / sec sustained ingest single-client · diurnal pattern · write-stalls rare under documented sizing. |

11 · HA & DR

# SURVIVE THE UNEXPECTED.

HA and DR are separate concerns. HA keeps the platform available when nodes or zones fail; DR restores operation when a region or dataset is lost. XERJ addresses them with different mechanisms — replication for HA, snapshots for DR.

## AVAILABILITY MODEL

|                      |                                                                                                                                        |
|----------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| REPLICATION          | Primary + N replicas per index · sync for the HA cluster topology · async for cross-region<br>ROAD Q3. Quorum writes.                  |
| FAILURE DETECTION    | Gossip-based health · bounded-latency failover · automatic replica promotion · control-plane notification.                             |
| READ-PATH CONTINUITY | Lock-free reader view — a failover swaps the pointer · in-flight queries complete on the stale view · new queries hit the new primary. |

## DISASTER RECOVERY

|                 |                                                                                                                                          |
|-----------------|------------------------------------------------------------------------------------------------------------------------------------------|
| SNAPSHOTS       | Segment-level snapshots to S3 / GCS / Azure Blob. Incremental (only new segments). Encrypted with customer KMS keys (BYOK Q2).           |
| RESTORE         | Point-in-time restore at segment granularity. Replay WAL from snapshot to exact seq_no. Warm standby mounts snapshots for near-zero RTO. |
| CROSS-REGION DR | Async replication of primary segments to a standby region. DNS-based failover.<br>ROAD Q3 2026.                                          |

## RTO / RPO TARGETS BY TOPOLOGY

| TOPOLOGY                   | RTO     | RPO        | AVAILABILITY |
|----------------------------|---------|------------|--------------|
| Single-node                | minutes | fsync      | best effort  |
| HA cluster · 3-node        | < 30 s  | 0 (sync)   | 99.95 %      |
| Multi-region active-active | < 2 min | sub-second | 99.99 %      |
| Air-gapped · 3-node        | < 30 s  | 0 (sync)   | 99.95 %      |

12 · REFERENCE ARCHITECTURES

# FOUR BLUEPRINTS. ONE PLATFORM.

Each blueprint is a validated topology with named upstreams, named consumers, retention policy, and sizing guidance. Every blueprint uses the same XERJ binary — what changes is the ingest surface, retention, and connected tools.

01 · SIEM MODERNIZATION

CISO · SOC

Firewall · endpoint · auth logs → Vector/Filebeat → XERJ (ES \_bulk)  
→ Kibana dashboards · alerting via explain-plan triggers  
retention 30 d hot · 90 d warm · 1 y cold  
sizing: 3-node · ~50 M events/node/day · disk ~2.5 TB/node

02 · RAG AT ENTERPRISE SCALE

CDO · CAIO

Documents → embedder (Voyage · OpenAI) → XERJ hybrid index  
→ LangChain / LlamaIndex retriever → application  
quantization SQ8 default · float32 re-rank for top-K  
sizing: ~55 M × 768-dim embeddings/node · hybrid p95 < 40 ms

03 · UNIFIED OBSERVABILITY

SRE · PLATFORM

App + infra → Otel Collector → XERJ (OTLP beta)  
→ Grafana via Prometheus remote\_read · alerting  
retention: metrics 30 d raw · logs 14 d hot · traces 7 d  
sizing: 3-node · ~80 K events/sec sustained · 6 → 1 stack

04 · AGENT MEMORY PLATFORM

AI PLATFORM

Agent runtime → XERJ session indices (time-decay scoring)  
→ native hybrid retrieval · token-budget-aware  
retention: 90 d · tenant-isolated per team  
sizing: session-scoped indices · linear with active-agent count

EACH BLUEPRINT · NAMED UPSTREAMS · DOCUMENTED RETENTION · SIZING TABLE

13 · COMPARISON

# WHERE THE CONSOLIDATION COMES FROM.

Three stacks an architect typically inherits. The XERJ column lists what one deployment of XERJ replaces — not what it adds on top.

## XERJ VS ES + PINECONE + REDIS (RAG STACK)

| AXIS                             | XERJ                  | LEGACY RAG STACK                          |
|----------------------------------|-----------------------|-------------------------------------------|
| SYSTEMS                          | 1 binary              | 3 systems + orchestrator                  |
| HYBRID RETRIEVAL                 | Native RRF, one query | Python orchestrator merges two HTTP calls |
| VECTOR MEMORY · 10 M × 1.5 K DIM | 18 GB (SQ8)           | 92 GB (float32)                           |
| VENDOR CONTRACTS                 | One                   | Three + framework                         |

## XERJ VS ELASTIC STACK (ES + KIBANA + LOGSTASH)

| AXIS                 | XERJ                  | ELASTIC STACK 8.13 |
|----------------------|-----------------------|--------------------|
| INGEST BINARIES      | 1 (built-in _bulk)    | 2 (ES + Logstash)  |
| IDLE CLUSTER MEMORY  | 400 MB · 1 node       | 8.5 GB · 4-node    |
| COLD START           | 50 ms                 | ~15 s JVM warm-up  |
| CLIENT COMPATIBILITY | ES 8.13 wire-protocol | ES 8.13 native     |

## XERJ VS SPLUNK + PROM + TEMPO (OBSERVABILITY)

| AXIS               | XERJ                     | SPLUNK + PROM + TEMPO      |
|--------------------|--------------------------|----------------------------|
| BINARIES           | 1                        | 6                          |
| QUERY LANGUAGES    | 1 (ES DSL + SQL roadmap) | 3 (SPL + PromQL + TraceQL) |
| MONTHLY BILL LINES | 1                        | 5                          |
| RETENTION POLICIES | 1 unified                | 3 separate                 |

14 · OPERATIONS

# DAY TWO. BOARD-READABLE.

MODE REQUIREMENTS

| PROFILE           | CPU     | MEMORY | DISK        | NETWORK |
|-------------------|---------|--------|-------------|---------|
| Small · pilot     | 4 vCPU  | 16 GB  | 500 GB NVMe | 1 Gbps  |
| Medium · prod     | 8 vCPU  | 64 GB  | 2 TB NVMe   | 10 Gbps |
| Large · scale-out | 16 vCPU | 128 GB | 8 TB NVMe   | 25 Gbps |

OBSERVABILITY OUT OF THE BOX

|                       |                                                                                                                                                             |
|-----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| METRICS               | Prometheus on /metrics. 13 counters / histograms / gauges across ingest, query, flush, cache. Per-instance registry — no global-singleton label collisions. |
| HEALTH /<br>READINESS | GET /healthz for liveness. GET /readyz holds until WAL replay completes — rolling restart fleet-safe by default.                                            |
| LOGS                  | Structured JSON to stdout. One event per line. Correlation ID on every request. Tail with jq, ship with any log-forwarding agent.                           |
| UPGRADES              | Rolling restart supported. Segment and WAL formats are forward-compatible across minor versions. One-minor-version window for rollback.                     |
| CONFIG                | Single TOML file · XERJ_CONFIG env · hot-reload via SIGHUP on non-structural changes. Config schema validated on load.                                      |
| BACKUP ·<br>RESTORE   | Snapshots to object storage. Incremental, encrypted, restorable at segment granularity. xerj snapshot create list restore.                                  |

## 15 · LICENSING &amp; SUPPORT

# THREE TIERS. ONE ENGINE.

## TIER 01

## COMMUNITY

- Full feature set · single-node
- Community support · GitHub · forum
- No SLA
- Ideal for POCs, dev, OSS projects

## TIER 02

## PROFESSIONAL

- HA cluster · 3 nodes
- Business-hours support · 24 h response
- 99.95 % availability SLA
- Quarterly health review
- Certified releases only

## TIER 03

## ENTERPRISE

- Multi-region · air-gapped · BYOK
- 24×7 support · 1 h response
- 99.99 % availability SLA
- Dedicated CSM + solutions architect
- Private roadmap input · early-access builds
- HIPAA BAA · custom compliance

### ROADMAP · 12 MONTHS

**Q2****2026  
FOUNDATION**

OIDC + SAML · RBAC · Helm chart · Terraform module · syslog ingest · BYOK encryption-at-rest

**Q3****2026  
SCALE**

Sharded ingest · cross-region async replication · Prometheus remote\_write ingest · Kafka source · SOC 2 Type I

**Q4****2026  
SURFACE**

SQL over HTTP · S3/GCS batch import · OTLP trace support · ISO 27001 · ML-ranking hooks

**'27****2027  
MANAGED**

Hosted SaaS offering · region-pinned · BYOK standard · HIPAA BAA · FedRAMP Moderate pursuit

## 16 · GET STARTED

# PROVE IT ON YOUR DATA.

The fastest path is a two-week pilot on a real workload. We send the binary and the reproduction scripts for every number in this brief. You send one SIEM / RAG / observability workload. We meet again in two weeks with your numbers, not ours.

## CONTACT

# HELLO@XERJ.AI

XERJ.AI · XERJ.COM · XERJ.DEV

## PILOT PROCESS

|        |                                                                                                                                   |
|--------|-----------------------------------------------------------------------------------------------------------------------------------|
| WEEK 0 | Scoping call · pick workload · sign mutual NDA if needed (not required to receive binary).                                        |
| WEEK 1 | Deploy binary in your environment · reproduce our battle reports on sample data · smoke-test ES client compatibility.             |
| WEEK 2 | Shadow-deploy against your workload · compare latency, cost, memory. Results meeting: your numbers, your hardware, your decision. |