

ImageInsight - Knowledge Base (KB-04)

ImageInsight **Knowledge Base 04** advances scalability, governance, and developer extensibility with 50 targeted enhancements spanning algorithms, runtime, security, compliance, observability, and platform integrations, building on patterns established in earlier releases while remaining backward compatible at the API and configuration levels.

Multimodal Feature Fusion

- Label: Fusion Strategy
- Field: similarity.fusion.strategy
- Configuration: weighted|late|co-attention (default: weighted); per-project tunable weights 0.0-1.0
- Details: Combines keypoint, perceptual, and text embeddings for improved match stability in noisy, mixed-content inputs.

Adaptive Threshold Curves

- Label: Threshold Mode
- Field: similarity.threshold.mode
- Configuration: static adaptive percentile (default: static); adaptive learns per-dataset ROC cutoffs nightly
- Details: Reduces false positives across heterogeneous datasets through data-driven dynamic thresholds.

SuperPoint v2 Upgrade

- Label: Keypoint Model
- Field: similarity.keypoint.model
- Configuration: orb superpoint superpoint_v2 (default: superpoint_v2)
- Details: Improves repeatability under motion blur and low light, with SIMD-optimized descriptors.

FSIM+ Extension

- Label: Similarity Metric
- Field: similarity.metric
- Configuration: ssim|fsim|fsim_plus (default: fsim_plus)
- Details: Adds gradient-phase coupling for better sensitivity to structural changes than FSIM alone.

Multi-GPU Node Affinity

- Label: Acceleration Affinity
- Field: runtime.acceleration.affinity
- Configuration: none socket|numa device (default: device)
- Details: Places workloads on optimal GPUs for reduced PCIe hops and improved throughput.

GPU Pool Auto-Scaler v2

- Label: Acceleration Mode
- Field: runtime.acceleration
- Configuration: auto manual cluster_auto (default: cluster_auto) with predictive scaling
- Details: Forecast-driven scaling trims tail latency under spiky demand while controlling cost.

Pipeline Plugin Sandbox

- Label: Plugin Isolation
- Field: preprocess.pipeline.plugins.sandbox
- Configuration: off|gvisor|firecracker (default: gvisor)
- Details: Executes custom preprocessing modules in isolated sandboxes with resource caps.

Plugin Schema Linter

- Label: Plugin Validation
- Field: preprocess.pipeline.plugins.schema.strict

- Configuration: true false (default: true)
- Details: Enforces semantic validation against JSON Schema with versioned contracts.

Region-Aware Redaction 2.0

- Label: Redaction Policy
- Field: compliance.redaction.policy
- Configuration: adds polygon masks and language-aware OCR regex packs
- Details: Extends redaction to multi-language PII with per-region templates and auditability.

OCR-Driven Rules Library

- Label: Redaction Ruleset
- Field: compliance.redaction.ruleset
- Configuration: default|pci hipaa custom; composable rule includes
- Details: Ships curated profiles for credit cards, MRNs, and IDs, customizable per tenant.

Tiered Retention Classes

- Label: Retention Class
- Field: governance.retention.class
- Configuration: short(7d) | standard (90d) | long (365d) | custom; purge-on-expiry toggle retained
- Details: Simplifies lifecycle management with preset classes and cost-optimized archival.

Legal Hold Overrides

- Label: Legal Hold
- Field: governance.retention.legal_hold
- Configuration: true false; scoped by case_id and tags
- Details: Overrides deletion for items under investigation with full audit linkage.

SIEM Field Mapping Packs

- Label: Audit Mapping
- Field: governance.audit.mapping
- Configuration: splunk|qradar|datadog|custom (default: splunk)
- Details: Standardizes audit JSON fields for drop-in SIEM ingestion without custom parsers.

Audit Integrity Seals

- Label: Audit Integrity
- Field: governance.audit.integrity
- Configuration: hmac|sha256|sha3 (default: sha256)
- Details: Adds per-record digests and chain links to prevent tampering.

Syslog TLS and MTLS

- Label: Audit Transport
- Field: governance.audit.transport
- Configuration: syslog+tls|json+https|kafka+mtls
- Details: Secures audit shipping with certificate rotation and mTLS options.

Project-Level Threshold Profiles

- Label: Threshold Profile
- Field: similarity.threshold.profile
- Configuration: default|strict|lenient custom with named profiles
- Details: Enables per-project sensitivity tuning for business-specific risk tolerances.

Dataset Drift Alerts

- Label: Drift Detection
- Field: observability.drift.enable
- Configuration: true false (default: true) with KS and PSI tests

- Details: Monitors distribution shifts in inputs and raises alerts with suggested retuning.

Canary Similarity Tests

- Label: Canary Tests
- Field: `quality.canary.enable`
- Configuration: `true false`; `percent_traffic`: 0.1-10
- Details: Routes a slice of traffic to new metrics for safe rollout and auto-rollback.

Batch Job Checkpointing

- Label: Batch Checkpoints
- Field: `runtime.batch.checkpoint.interval`
- Configuration: N items or minutes; resumable runs
- Details: Protects long-running comparisons against node failures and preemptions.

Priority Queues

- Label: Job Priority
- Field: `runtime.queue.priority`
- Configuration: `low|normal|high` critical with fair-share per tenant
- Details: Ensures urgent workloads preempt capacity without starving lower tiers.

Request Throttling Policies

- Label: Throttle Policy
- Field: `api.throttle.policy`
- Configuration: `per_key|per_ip|per_project` limits; burst and sustained
- Details: Prevents abuse while preserving burst capacity for legitimate clients.

OpenSearch Vector Sync

- Label: Index Sync
- Field: `index.vector.sync`

- Configuration: realtime interval manual (default: interval)
- Details: Keeps vector indices consistent across shards and regions for RAG use cases.

Pinecone Connector

- Label: External Vector Store
- Field: index.vector.external
- Configuration: pinecone opensearch|aurora_pgvector (multi)
- Details: Adds pluggable vector store abstraction with retryable writes and schema checks.

Aurora pgvector Optimizations

- Label: Vector Tuning
- Field: index.vector.pg.settings
- Configuration: ivfflat probes, HNSW M and ef_search per-collection
- Details: Improves recall/latency trade-offs for serverless vector workloads.

Cross-Region Replication

- Label: Replication Mode
- Field: storage.replication.mode
- Configuration: off async sync (RPO 0/near-0)
- Details: Enables disaster recovery and low-latency access for global teams.

Content Addressable Store

- Label: CAS Enable
- Field: storage.cas.enabled
- Configuration: true false (default: true)
- Details: Deduplicates identical artifacts with SHA-256 keys to lower storage cost.

Encryption Key Rotation

- Label: KMS Rotation
- Field: security.kms.rotation.days
- Configuration: 30-365 (default: 180)
- Details: Enforces periodic CMK rotation with auto-rewrap of stored items.

Fine-Grained RBAC

- Label: Role Policy
- Field: security.rbac.policy
- Configuration: viewer|analyst|operator|admin|custom
- Details: Scopes permissions to projects, data classes, and actions with policy-as-code.

SSO Claims Mapping

- Label: SSO Mapping
- Field: security.sso.claims.map
- Configuration: map OIDC/SAML attributes to roles and projects
- Details: Automates least-privilege assignments from identity provider attributes.

Secrets Mounts for Plugins

- Label: Plugin Secrets
- Field: preprocess.pipeline.plugins.secrets
- Configuration: vault ssm|env (scoped)
- Details: Supplies short-lived credentials to plugin sandboxes securely.

Lineage Propagation

- Label: Lineage Enable
- Field: governance.lineage.enabled
- Configuration: true|false (default: true)
- Details: Tracks provenance from input through preprocessing, metrics, and decisions.

Decision Policy Engine

- Label: Decision Policy
- Field: decision.policy.dsl
- Configuration: CEL-like expressions with versioning and tests
- Details: Encodes accept/review/reject logic with auditable, testable rules.

Explainability Traces

- Label: XAI Trace
- Field: observability.xai.trace
- Configuration: off minimal full (default: minimal)
- Details: Captures intermediate scores, thresholds, and feature attributions for reviews.

Cost Attribution Tags

- Label: Cost Tags
- Field: billing.tags.required
- Configuration: true false; required keys: project, env, owner
- Details: Enables granular showback/chargeback on compute, storage, and indexing.

API Idempotency Keys

- Label: Idempotency
- Field: api.idempotency.enable
- Configuration: true false (default: true)
- Details: Prevents duplicate submissions under network retries using client-supplied keys.

Request/Response Schemas

- Label: API Schema
- Field: api.schema.validate

- Configuration: strict warn|off (default: strict)
- Details: Validates REST and Webhook payloads with clear errors and versioned schemas.

Webhook Retries + DLQ

- Label: Webhook Policy
- Field: integration.webhook.retry
- Configuration: exponential backoff, max_attempts, dead_letter.topic
- Details: Improves reliability for downstream notifications and automations.

Kafka Event Streams

- Label: Event Stream
- Field: integration.events.kafka
- Configuration: topic mappings for jobs, results, audits
- Details: Emits structured events for real-time analytics and pipelines.

S3 Multipart Optimization

- Label: Upload Strategy
- Field: storage.upload.multipart.min_part_mb
- Configuration: 5-64 (default: 8)
- Details: Tunes part size and concurrency to speed large artifact uploads.

Checksum Verification

- Label: Upload Checks
- Field: storage.upload.verify_checksum
- Configuration: true false (default: true)
- Details: Ensures end-to-end integrity with MD5/SHA verification on ingest.

Preprocessing Graph DSL

- Label: Pipeline Graph

- Field: preprocess.pipeline.graph
- Configuration: YAML DSL supporting branches, loops, and gating
- Details: Enables complex pipelines beyond linear plugins with validations and dry runs.

Parallel Stage Execution

- Label: Parallelism
- Field: preprocess.pipeline.parallelism
- Configuration: auto|N workers per stage
- Details: Cuts latency by concurrently executing independent preprocessing steps.

Media Type Adapters

- Label: Modality Adapter
- Field: preprocess.adapters.enable
- Configuration: image|pdf|video|audio (multi-select)
- Details: Normalizes diverse media into consistent feature tensors for downstream metrics.

Temporal Consistency Checker

- Label: Temporal Check
- Field: similarity.temporal.consistency
- Configuration: off|light|strict (default: light)
- Details: Penalizes frame-to-frame anomalies to reduce flicker in video comparisons.

Heatmap Exports

- Label: Visualization
- Field: output.visual.heatmap
- Configuration: off|png|json (mask)
- Details: Produces interpretable similarity maps for QA and stakeholder reviews.

Result Templating

- Label: Output Template
- Field: output.template.id
- Configuration: default|minimal|audit|custom (handlebars-like)

Scheduled Policy Changes

- Label: Policy Schedule
- Field: governance.policy.schedule
- Configuration: cron expressions with preview and conflict checks
- Details: Automates after-hours policy updates with rollbacks on errors.

Tenant Quotas

- Label: Quota Policy
- Field: tenancy.quota.limits
- Configuration: requests/min, storage_gb, gpu_hours per tenant
- Details: Protects shared clusters and enables fair use agreements.

Data Residency Controls

- Label: Residency Region
- Field: compliance.data.residency
- Configuration: region-locked projects with block/allow lists
- Details: Ensures datasets remain in approved geographies for regulatory compliance.

Enhanced Audit Logging v2

- Label: Audit Enable
- Field: governance.audit.enabled
- Configuration: true|false (default: true); enriched context fields and redaction flags
- Details: Adds context-rich entries, mapping to SIEM packs with integrity seals for end-to-end auditability.

