

CEDAR 4.0 → QUASAR

Architecture Review / Feedback Session



Adriel Casellas, Allan Almeida | Jun 9, 2025

Quasar

Next Generation Cedar

- AI-enhanced, cloud-native Document Management System
- Replaces Cedar 3.0 with immutable, encrypted, chunked storage
- Built for legal, court, and regulated environments
- Supports both eFlex (short-term) and Alpine (long-term) lifecycles

Core Concepts

Block / Blob / View

- **Block:** Small encrypted chunk of binary data (e.g. ~120KB - 1MB depending on content type)
- **Blob (File):** Sequence of blocks forming a full document
- **View:** Virtual transformation of a Blob (e.g., stamped, OCR'd) without data duplication
- Think Dropbox meets Git. Views give flexibility for derived documents without copying.

Smart Uploads

Chunked CAS + Encryption at rest

- Content-Addressable Storage (CAS): Each chunk hashed (BLAKE3 or SHA256/SHA1)
- Chunks stored and referenced immutably
- Smart chunking for PDFs:
 - Split by object streams
 - Decompress object streams before hashing
 - Enables finer diffing and reuse
- Reduces reprocessing costs and speeds up OCR/stamping. Foundation for deduplication.

Encryption Model

- Each block is encrypted with a derived key:
 - Requires **master key** (server-side)
 - And **original content hash** (known only if unaltered)
- Even if a blob leaks, it is unreadable without the key + hash
- Encryption at rest: Per-org / per-file keys
- We sell this as “zero trust” content security — content and metadata both protected.

Memory + Storage Efficiency

How We Sell Efficient Storage

- Deduplication across:
 - Versioned uploads
 - Shared filings across matters
 - OCR'd or transformed views
- **Memory savings:** Only active chunks streamed in/out
- Estimated disk savings:
 - 30–60% in typical usage
 - Up to 90% with shared filings (e.g., class actions)

FIPS-Ready Cryptography

How We Sell to Security Conscious Clients

- FIPS (Federal Information Processing Standards) compliant means a product, system, or solution adheres to the security requirements outlined in a specific FIPS publication.
- Default: BLAKE3 (non-FIPS, fast)
- FIPS-compliant: SHA256, AES-256, routed through AWS KMS or local HSM
- Deployment mode: Toggleable at runtime or build time
- Crucial for federal and state court contracts. Meets CJIS / HIPAA when needed.

Metadata and Search

How We Sell Document Insights

- Rich namespaced metadata (JSON-based)
- PostgreSQL with:
 - Full-text search
 - Vector embedding search (via PGAI)
 - Audit trails and version patch history
- One DB for both metadata and semantic/AI search makes ops easier

Access Control

How We Sell Trust

- Clearance levels (e.g., public, restricted, private)
- Hybrid RBAC + ABAC:
 - Users/roles
 - Scoped rules (per org/matter/doc)
- API and SQL enforceable
 - JWT tokens with scoped claims
- This goes beyond typical DMS. ABAC can enforce clearance + org policy.

Upload + Job Pipeline

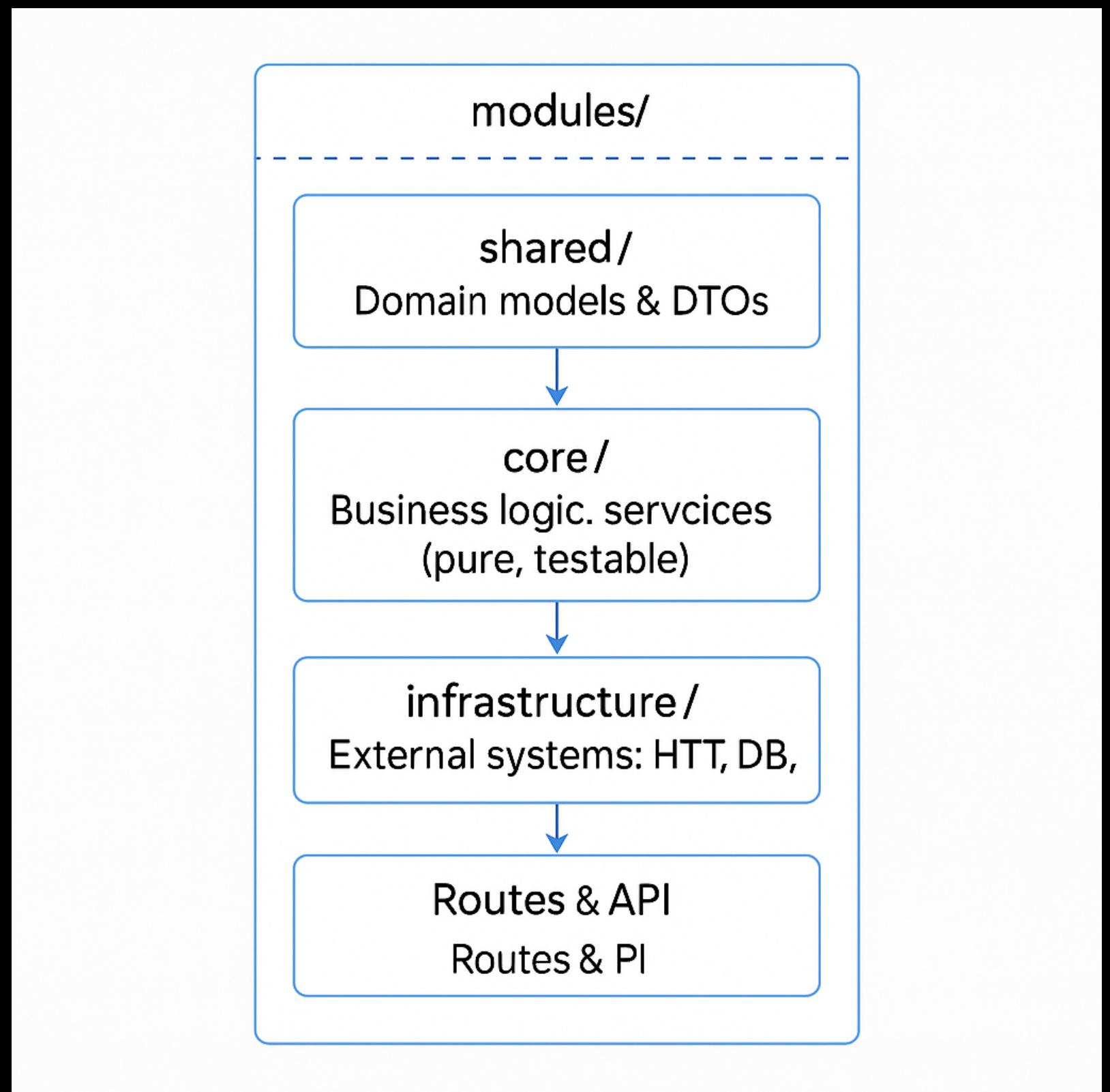
How We Sell Reliability

- Streaming Insert API:
 - Methods include: insertBlock, insertBlob, insertBatch (app level, not exposed to users)
 - All writes are validated (MIME, size, integrity)
 - Guards against 1TB zip bomb attacks :)
 - Optional plugin chain on insert:
 - OCR → Classify → Stamp
 - Jobs can run async or inline; plugins can trigger based on content type.

Plugin Architecture

How We Sell Plugins (and Custom Extensions)

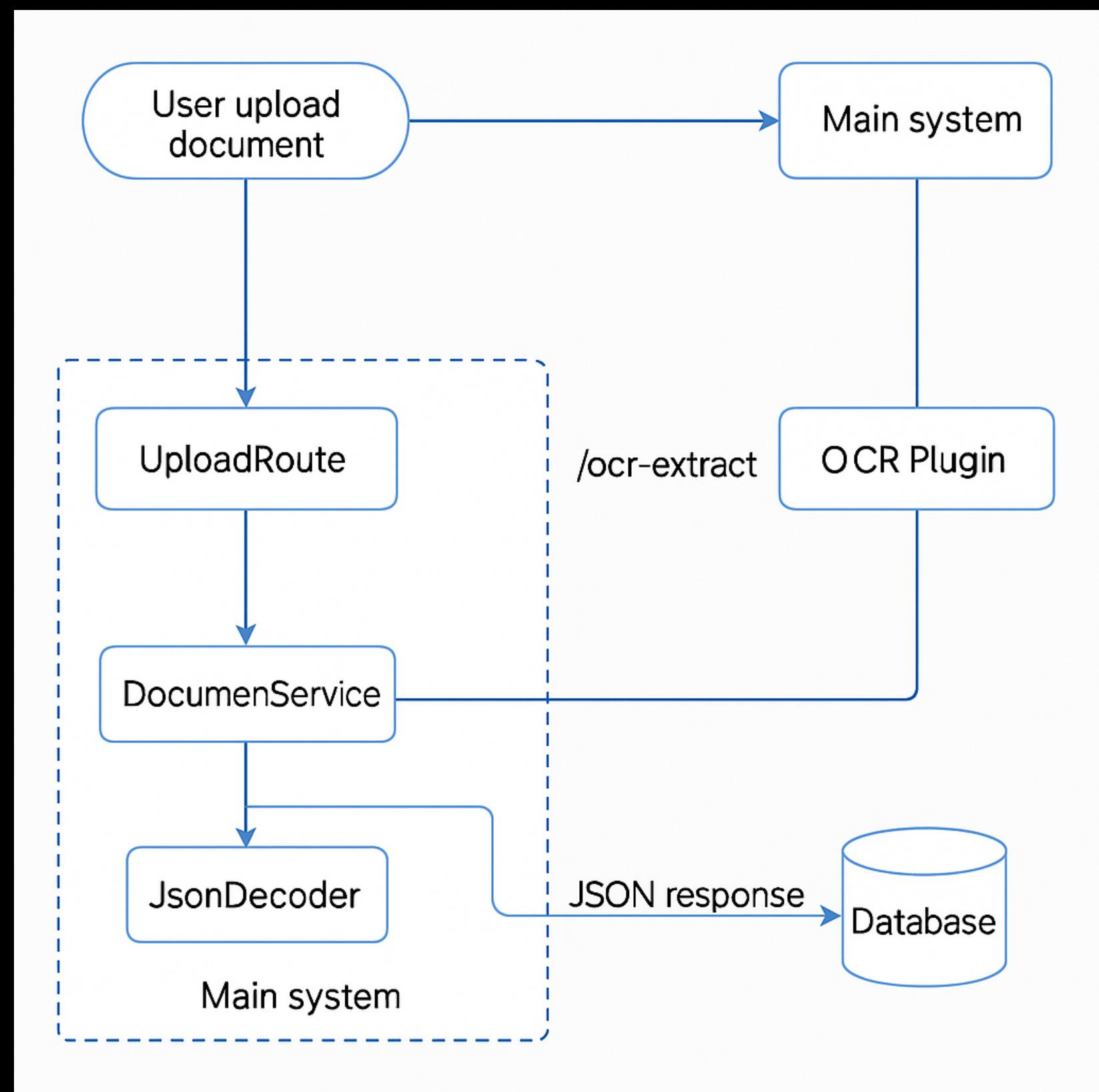
- Plugin = Declarative, isolated job
- Built-in: OCR, PDF/A, tagging, summarization, virus scan
- Language/Tech Agnostic, Docker execution
- Can modify metadata, emit new view, or enrich AI search index
- Plugins make Quasar composable. Add-ons without redeploys.



Example: OCR Plugin Data Flow

Similar Dataflow for other Plugins as well

- User uploads a document through the application interface.
- Request is routed via `UploadRoute` to internal document processing logic
- `DocumentService` orchestrates the flow:
 - Prepares document metadata or file path.
 - Sends HTTP request to `/ocr-extract` endpoint exposed by the OCR plugin.
- OCR plugin processes the document and extracts text
- Plugin returns a structured **JSON response** containing the extracted content.
- ZIO-Schema Json Decoder parses the response.
- Parsed data is saved into the **Database** for further use.



WebDAV Integration (TBD)

How We Integrate with eFlex/Alpine

- Read-only WebDAV layer:
 - Virtual filesystem → Views
- Use cases:
 - Court staff browsing in Alpine
 - Legacy app compatibility
- Fully audit-logged, token-scoped
- Easy way to bridge into e.g., Windows Explorer or other legal tools (eFlex/Alpine/etc).

System Deployment

How We Sell Infrastructure

- Cloud-native and on-prem parity
 - AWS S3 for cloud, MinIO for on-prem FS, same API for both
- Supports:
 - Docker Swarm / Kubernetes
 - PostgreSQL, S3 or MinIO
 - Redis (short-term cache)
- Plugin runners use sidecar/worker pattern
- Great for hybrid deployments. No lock-in, runs air-gapped.

Feature Summary

What's Different?

Feature	CEDAR 3.0	Quasar
SOAP API	✓	🚫 (REST/gRPC)
CAS + Deduplication	✗	✓
Encrypted Blocks	✗	✓
Versioning	Partial	Full (via Views)
OCR/Tagging	✗	✓
Namespaced Metadata	✗	✓
AI Integration (LLM)	✗	✓
WebDAV	✗	✓

AI Enhancements

How We Sell AI to Courts

- OCR + Summarization + Entity Tagging via LLMs (Ollama)
- Semantic search (RAG)
- Future: Predictive doc categorization, live form generation
- Cedar's AI layer: No need to bolt on another platform.