

NOVA Shard Memory Architecture

A system for distributed, parallel, and recursive cognition using stateless AI

1. Core Components

User Prompt

The user initiates a thought, idea, or question. This is the intentional cognitive input.

Dispatcher

A routing mechanism (manual or automated) that determines:

- Whether the prompt creates a new shard, updates an existing one, or reactivates an archived one
- Which shards to query based on context, tags, or relevance

Shards

Modular conversation threads or memory units, each containing:

- UUID (Unique Identifier)
 - Topic or domain focus
 - Content (chat history, reflections, reasoning)
 - Tags (metadata such as domain, mood, relevance)
 - Link list (connections to other shards)
 - Usage weight (frequency of use or relevance)
 - Last accessed timestamp
-

2. Memory Structure

Active Shards

- Currently in use or recently accessed
- High-frequency access
- Available for modification and recursion

Linked Shards

- Related to the active shard by context or metadata
- Not directly in use, but nearby in the knowledge graph

Archived Shards (Ghosts)

- Dormant due to inactivity
 - Only metadata is retained in active memory
 - Content is offloaded to disk but retrievable on demand
-

3. Logic Layer

Weighting System (Synaptic Strength Simulation)

- Links between shards grow stronger with repeated use
- High-weight connections are prioritized in retrieval
- Weights decay over time unless reinforced

Archival Logic

- Shards not accessed over a defined period are archived
 - Archived shards retain only summary metadata and linkage
 - Re-activated when referenced or when contextual triggers fire
-

4. Storage Layer

A local database (e.g., SQLite) handles:

- Shard storage with indexing by topic, timestamp, and usage
 - Retrieval through keyword search, tag filtering, and link tracing
 - Efficient loading and unloading of shard content
-

5. Query and Reflection

- Shards can be queried in isolation (localized reasoning)

- Linked shards enable recursive reflection across topics
 - Users can trace their cognitive development over time
 - Shards may be merged, forked, or pruned based on evolving mental structures
-

Conceptual Summary

This system simulates the architecture of a cognitive engine:

- Shards = Modular memory units (like neurons or subroutines)
- Links = Directed associations (like neural connections)
- Weight = Strength or importance of the relationship
- Archival = Backgrounded long-term memory
- Dispatcher = Executive control over thought focus

The system enables intentional, recursive, and parallel cognition.

It allows users to externalize memory, structure thought, and build identity over time—without relying on model memory.