SHARDS: Semantic Clustering

Shards are semantic grouping values that can link memories together.

Shards will need a standardized shard spec

Things like name, timestamp, confidence (used for RAG)

🧠 What's the Difference Between Embeddings and Shards?

> **Embeddings** are raw **coordinates in semantic space.**
> **Shards** are **named or meaningful clusters** derived *from* those coordinates.

📍 Embeddings = the data

Embeddings are raw coordinates

- Each memory has one
- Represents that memory's **semantic position** in high-dimensional space

🧩 Shards = interpreted structure

- Group **related** memory entries based on closeness in embedding space (and optionally time, tags, emotion, etc.)
- They turn *spatial patterns* into *semantic threads*

Shards are neighborhood names

Think of it like:

- Embeddings: every star in the sky has X/Y/Z
- Shards: constellations you draw from how those stars cluster together

Embeddings cant capture: time-based context, emotional tags, relational context (such as chats, who said it, which persona was active)

🧩 Suggested Shard Types

Here's a proposed shard taxonomy:

Shard Type	What It Represents	How It's Generated
Semantic Shard	Cluster of conceptually similar memories	Embedding clustering (e.g., UMAP + HDBSCAN)
Temporal Shard	Burst of activity over time	Time window + tag co-occurrence
Emotional Thread	Shared emotional tags / tone	Tags or emotion classifier
Conversational Shard	Memories from the same chat or session	Conversation ID / dialogue ID
Persona Shard	Related to a specific Prism/Humor config	Prism ID or persona state at time of memory
Event-Based Shard	User-labeled shard (e.g., "Breakup Week")	User-defined + auto-backfilled by time range

Each **memory entry** links to **multiple shard objects**, like:

json Copy Edit

```json
"shards": [
  {
    "id": "shard_4521",
    "type": "semantic",
    "name": "career_direction_conflict",
    "confidence": 0.88
  },
  {
    "id": "shard_7822",
    "type": "temporal",
    "name": "apr_2025_syncup_spike",
    "confidence": 1.0
  },
  {
    "id": "shard_9014",
    "type": "emotional",
    "name": "burnout_loop",
    "confidence": 0.92
  }
]
```

Each `shard` is stored in a separate `shards` collection, with its metadata, generation source, and linked memory IDs.