AN

NLTlabs

NEW LOGIC TECH

# NLT Memory

## Deep Dive · Cross-Session Recall

*HTTP bridge · tiered retention · hive · knowledge graph.*

44 tapps_memory actions on one MCP tool surface.
Consumers never embed tapps-brain in-process — bridge only.
Architectural tier persists decisions across Ralph loops and IDE sessions.
Outward volumes use NLT Memory — not maintainer checkout names.

**AUDIENCE** Engineering · Platform operators · Agent authors

**NLT LABS** NEW LOGIC TECH

Version · memory Issued · 2026-06-11

# Contents



**How to read this.** Generated from the NLT Quality Pipeline checkout at /home/wtthornton/code/NLT Quality Pipeline. Action tables mirror docs/MEMORY_REFERENCE.md. Outward copy uses **NLT Memory** — meta ops docs may still say NLT Memory for compose labels.

PART I

# NLT Memory at a glance

**NLT Memory** is the shared recall authority for every agent surface in the NLT portfolio. IDE agents, Ralph loops, and AgentForge workflow nodes reach it through the NLT Quality Pipeline memory bridge — never as an in-process Python embed in consumer repositories.

## What lives in NLT Quality Pipeline vs NLT Memory

**NLT Quality Pipeline** owns the MCP tool surface (tapps_memory), bridge wiring, hooks, and per-project .NLT Quality Pipeline.yaml overrides. **NLT Memory** owns Postgres, embeddings, tier decay, federation hub storage, and native graph tools. Consumer repos read sibling checkouts at PDF build time only.

| Layer | Owner | Runtime role |
|---|---|---|
| MCP tool | NLT Quality Pipeline | Single tool, action= dispatch (42 actions) |
| Bridge | tapps-core BrainBridge | HTTP or native MCP to brain |
| Store | NLT Memory | Postgres + embeddings + audit JSONL |
| Federation hub | ~/.NLT Quality Pipeline/memory/ | Cross-project federated.db |
| AgentForge invoke | BrainPool | recall_for_prompt during workflow nodes |

> “Same input always produces same output in the retrieval path — BM25 ranking, exponential decay, and graph walks are algorithmic. No LLM calls in memory recall.”
>
> **— Deterministic recall**

# Package map

Primary implementation anchors below. Citations in this module use path::symbol form for report-studio verify.

| Path | Role |
|---|---|
| packages/NLT Quality Pipeline/src/tapps_mcp/server_memory_tools.py | tapps_memory action dispatch + register |
| packages/tapps-core/src/tapps_core/brain_bridge.py | BrainBridge HTTP/native bridge to NLT Memory |
| packages/NLT Quality Pipeline/src/tapps_mcp/server_helpers.py | _get_brain_bridge, _get_memory_store |
| docs/MEMORY_REFERENCE.md | 42-action public reference (source for this volume) |
| /home/wtthornton/code/NLT Quality Pipeline/.NLT Quality Pipeline.yaml | Per-project memory.* and memory_hooks overrides |

## Part II — HTTP bridge

PART II

# BrainBridge call flow

Every recall and save crosses the same bridge boundary. Profile gating runs before the HTTP round trip; degraded payloads are structured JSON — not silent failures.

| Step | Component | Detail |
|---|---|---|
| 1 | IDE / agent | Calls tapps_memory(action=search, query=...) |
| 2 | server_memory_tools | tapps_memory dispatches to handler by action |
| 3 | BrainBridge | HTTP POST to NLT Memory /v1/tools or native MCP in dev |
| 4 | Profile gate | ToolNotInProfileError if action not in active profile |
| 5 | NLT Memory | memory_recall / memory_save / graph tools execute |
| 6 | Postgres + embeddings | Tier decay, FTS, optional vector rerank |
| 7 | Response | Structured JSON back through bridge to MCP client |
| 8 | Hooks (optional) | memory_hooks auto_recall / auto_capture on edit events |
| 9 | AgentForge | BrainPool.recall_for_prompt during workflow invoke (Vol 5) |
| 10 | Reinforce loop | reinforce / reinforce_many after useful recall |

# AgentForge runtime recall — BrainPool

Workflow nodes do not call tapps_memory directly during invoke. AgentForge BrainPool.recall_for_prompt batches recall for the active agent profile before model generation — see Vol 5 for orchestration context.

| Phase | Component | NLT Memory surface |
|---|---|---|
| Pre-invoke | BrainPool | recall_many / search scoped to brain_profile |
| Generate | Agent runtime | Prompt grounded with recalled entries |
| Post-invoke | Workflow node | Optional reinforce on useful findings |
| Cross-agent | Hive propagate | Shared namespace when Agent Teams on |

> “NLT Engine agents declare tapps-brain in AGENTS.md — the bridge enforces tier, profile, and content-safety gating. Consumers never embed Postgres drivers.”
>
> **— Consumer boundary**

# HTTP bridge vs in-process mode

Production portfolios use HTTP bridge to NLT Memory with TAPPS_BRAIN_AUTH_TOKEN. In-process native MCP suits local dev when the brain pool is co-located — graph tools require HTTP.

| Mode | When | Graph tools |
|---|---|---|
| HTTP | CI, AgentForge, remote brain | Full TAP-1630 surface |
| In-process | Local NLT Quality Pipeline doctor | knowledge_graph_requires_http_bridge |
| Degraded | Brain auth failure + tolerate flag | Structured errors, no swallow |

| Symptom | First check | Volume |
|---|---|---|
| brain_auth_failed | TAPPS_BRAIN_AUTH_TOKEN | Vol 8 auth table |
| Empty recall | brain_bridge_health.ok | tapps_session_start |
| ToolNotInProfileError | profile_switch / profile_info | Action families |
| Graph action blocked | Bridge mode = http | Part V graph table |

# Auth and health

Set TAPPS_BRAIN_AUTH_TOKEN for authenticated HTTP bridge mode. tapps_session_start returns a brain_bridge_health block: dsn_reachable, native_health_ok, and roll-up ok. Agents should surface bridge errors instead of swallowing them.

| Signal | Where | Meaning |
|---|---|---|
| brain_bridge_health.ok | tapps_session_start | Bridge reach + native self-check |
| ToolNotInProfileError | BrainBridge | Action not in active brain profile |
| knowledge_graph_requires_http_bridge | Graph actions | In-process bridge cannot run graph tools |
| degraded: True | gc / consolidate | Structured fallback, not hard failure |

**Part III — Data model**

PART III

# Tiers and decay

Tier selects half-life for exponential decay. Reinforce resets the decay clock and may boost confidence. Architectural saves with auto_supersede_architectural keep temporal chains via supersede — responses may include superseded_old_key.

| Tier | Half-life | Use for | Examples |
|---|---|---|---|
| architectural | 180 days | Stable, long-lived decisions | We use PostgreSQL; monorepo with three packages |
| pattern | 60 days | Coding conventions and style | Use structlog not print; models inherit BaseModel |
| procedural | 30 days | Workflows and step sequences | Deploy: build → test → push → tag |
| context | 14 days | Short-lived session facts | Refactoring auth module this sprint |

## Composite search ranking

search ranks with BM25 over FTS5, then blends confidence, recency, and access frequency. Optional embedding vectors rerank when present on entries — still deterministic, no LLM in the retrieval path.

| Signal | Weight | Meaning |
|---|---|---|
| BM25 relevance | 40% | Lexical match over indexed entry text |
| Confidence | 30% | Stored entry confidence after reinforce/decay |
| Recency | 15% | Exponential decay by tier half-life |
| Frequency | 15% | Access count from successful recalls |

Low-similarity or empty results may auto-emit feedback_gap when memory.feedback_auto_emit is true — feeds the brain flywheel for operators to fill with rate or new save entries.

## Scopes and federation

save accepts project, branch, and session — not shared. Shared scope is a propagation tier used by federate_publish, not a direct save scope.

| Scope | Visibility | Use for |
|---|---|---|
| project | All sessions in this project (default) | Architecture, patterns, decisions |
| branch | Only sessions on this git branch | Branch-specific WIP |
| session | Current session only (expires 7 days) | Temporary notes |
| shared | Federation-eligible (cross-project) | Reusable knowledge published to hub |

# Federation operator playbook

Cross-project handoff uses the hub at ~/.NLT Quality Pipeline/memory/federated.db. Local .NLT Quality Pipeline/memory/memory.db stays project-scoped; only shared-scope entries publish upstream.

| Step | Action | Outcome |
|---|---|---|
| 1 | federate_register | Project id + tags in hub registry |
| 2 | save (shared scope) | Entries eligible for publish |
| 3 | federate_publish | Push shared entries to hub |
| 4 | federate_subscribe | Pull other projects by tag/confidence |
| 5 | federate_sync | Merge subscribed memories locally |
| 6 | federate_search | Query local + federated with local boost |

# Memory hooks — auto-recall and auto-capture

Project .NLT Quality Pipeline.yaml can wire memory_hooks so IDE sessions recall before edits and capture architectural saves after quality signals. Hooks call the same bridge — they are not a second memory stack.

| Hook | Trigger | Typical action |
|---|---|---|
| auto_recall | Session start / prompt | search ranked queries from config |
| auto_capture | quick_check pass / gate pass | save pattern tier signals |
| pre_compact | Context compaction | Index session chunks before trim |
| session_end | Chat end | session_end summary → brain index |

Disable noisy hooks per repo via memory_hooks.auto_recall: false — Vol 10 covers maintainer tuning; this volume documents the recall authority.

# Feedback flywheel — rate and gap signals

TAP-1632 closes the loop: empty search emits feedback_gap; agents and operators score entries with rate (feedback_rate). Ralph coordinator debriefs call rate on entries actually used in the loop.

| Signal | Source | Operator response |
|---|---|---|
| feedback_gap | search miss / low similarity | save missing pattern |
| feedback_rate + | Agent debrief | reinforce relied keys |
| feedback_rate – | Hallucination trace | delete or supersede entry |
| reinforce_many | Batch after epic | Boost confidence headroom |

# Outward naming — NLT Memory vs maintainer labels

Investor and operator volumes say **NLT Memory**. Maintainer checkouts, compose service names, and MCP tool ids may still read NLT Quality Pipeline or NLT Memory — cite those in engineering tables, not in brief copy.

| Outward (use in briefs) | Maintainer (code / compose) | Role |
|---|---|---|
| NLT Memory | NLT Quality Pipeline | MCP tool + bridge wiring |
| NLT Memory store | NLT Memory | Postgres + embeddings |
| Memory bridge | BrainBridge | HTTP/native adapter |
| FUND proof | portfolio.nltlabs.ai | Public portfolio cards |

**Part IV — Action surface**

PART IV

# tapps_memory action families

**44 actions** across **15 families.** Overview below; each family follows on its own page with full parameter tables.

| Group | Actions | Theme |
|---|---|---|
| Core CRUD | save, save_bulk, get, list, delete | Every consumer write path funnels through tapps_memory save with tier and scope. Architectural saves may supersede prior… |
| Search and ranking | search | Recall is deterministic — BM25 composite scoring with confidence, recency, and frequency. No LLM in the retrieval path. |
| Intelligence and maintenance | reinforce, gc, contradictions, reseed | Decay clocks, contradiction surfacing, and operator maintenance hooks. |
| Consolidation (degraded) | consolidate, unconsolidate | Consolidation actions remain on the MCP surface for compatibility but return structured degraded payloads on NLT Memory… |
| Import and export | import, export | Portable snapshots for handoff, backup, and federation publish prep. |
| Federation (cross-project) | federate_register, federate_publish, federate_subscribe, federate_sync, federate_search, federate_status | Hub at ~/.NLT Quality Pipeline/memory/federated.db — local project DB stays project-scoped. |
| Session indexing and validation | index_session, search_sessions, session_end, validate, maintain | Native session memory (TAP-1633) replaces the legacy local session index merge. |
| Knowledge graph (TAP-1630) | related, relations, neighbors, explain_connection | Graph tools require HTTP bridge; in-process mode returns knowledge_graph_requires_http_bridge. |
| Batch ops (TAP-1631) | recall_many, reinforce_many | Single-round-trip wrappers around brain memory_*_many endpoints. |
| Feedback flywheel (TAP-1632) | rate | Closes the loop on search misses and explicit agent ratings. |
| Security (Epic M1) | safety_check, verify_integrity | Safety and integrity checks before trusting recalled content. |
| Profiles and lifecycle (Epic M2) | profile_info, profile_list, profile_switch | Brain capability profiles gate which bridge tools are callable. |
| Health and diagnostics | health | Surfaced in tapps_session_start brain_bridge_health and tapps doctor. |

| Group | Actions | Theme |
|---|---|---|
| Hive and agent teams (Epic M3) | hive_status, hive_search, hive_propagate, agent_register | Cross-agent propagation when CLAUDE_CODE_EXPERIMENTAL_AGENT_TEAMS is enabled. |
| Pipeline lifecycle (internal) | session_start_capture, session_end_consolidate | Phase-2 lifecycle hooks — not daily driver actions on the public catalog. |

# Core CRUD

Every consumer write path funnels through tapps_memory save with tier and scope. Architectural saves may supersede prior keys when auto_supersede_architectural is on.

| Action | Parameters | Description |
|---|---|---|
| save | key, value, tier, scope, tags, source | Save entry; architectural tier may supersede |
| save_bulk | entries (list, max 50) | Batch save in one bridge round trip when HTTP |
| get | key | Retrieve by key with provenance for consolidated entries |
| list | scope, tier, tags, limit | Filtered list (max 50) |
| delete | key | Delete by key |

# Search and ranking

Recall is deterministic — BM25 composite scoring with confidence, recency, and frequency. No LLM in the retrieval path.

| Action | Parameters | Description |
|---|---|---|
| search | query, ranked, limit, scope, tier, tags | BM25 composite: 40% relevance + 30% confidence + 15% recency + 15% frequency |

**Note.** Empty or low-similarity search may auto-emit feedback_gap when feedback_auto_emit is enabled.

# Intelligence and maintenance

Decay clocks, contradiction surfacing, and operator maintenance hooks.

| Action | Parameters | Description |
|---|---|---|
| reinforce | key, boost | Reset decay; optional confidence boost (max +0.2) |
| gc | — | Archive stale memories via maintenance_gc (operator profile) |
| contradictions | — | Detect memories contradicting current project state |
| reseed | — | Re-seed from project profile without overwriting human entries |

# Consolidation (degraded)

Consolidation actions remain on the MCP surface for compatibility but return structured degraded payloads on NLT Memory 3.10+.

| Action | Parameters | Description |
|---|---|---|
| consolidate | entry_ids or query, dry_run | Degraded — memory_consolidate removed on brain |
| unconsolidate | key | Same degraded status as consolidate |

**Note.** Plan: deprecate consolidate/unconsolidate once deprecation cycle completes.

# Import and export

Portable snapshots for handoff, backup, and federation publish prep.

| Action | Parameters | Description |
|---|---|---|
| import | file_path, overwrite | Import JSON (max 500 entries) |
| export | file_path, format | Export JSON or Markdown |

# Federation (cross-project)

Hub at ~/.NLT Quality Pipeline/memory/federated.db — local project DB stays project-scoped.

| Action | Parameters | Description |
|---|---|---|
| federate_register | project_id, tags | Register project in federation hub |
| federate_publish | — | Publish shared-scope entries to hub |
| federate_subscribe | sources, tag_filter, min_confidence | Subscribe to other projects |
| federate_sync | — | Pull subscribed memories |
| federate_search | query | Search local + federated with local boost |
| federate_status | — | Hub status: projects, subscriptions, counts |

# Session indexing and validation

Native session memory (TAP-1633) replaces the legacy local session index merge.

| Action | Parameters | Description |
|---|---|---|
| index_session | session_id, chunks (JSON) | Index transcript chunks via memory_index_session |
| search_sessions | query, limit | Search indexed session chunks |
| session_end | value, tags, dry_run | Session-end summary via tapps_brain_session_end |
| validate | key | Validate entry against current project state (async) |
| maintain | — | Full maintenance cycle: gc + contradictions + reseed |

# Knowledge graph (TAP-1630)

Graph tools require HTTP bridge; in-process mode returns knowledge_graph_requires_http_bridge.

| Action | Parameters | Description |
|---|---|---|
| related | key, max_hops | Walk graph from key → memory_find_related |
| relations | key OR subject/predicate/object_entity | Entry relations or SPO triple query |
| neighbors | entry_ids, max_hops, limit, predicate | k-hop neighborhood |
| explain_connection | subject, object_entity, max_hops | Path explanation between entities |

# Batch ops (TAP-1631)

Single-round-trip wrappers around brain memory_*_many endpoints.

| Action | Parameters | Description |
|---|---|---|
| recall_many | entries (JSON query strings) | Batch recall via memory_recall_many |
| reinforce_many | entries ({key, confidence_boost?}) | Batch reinforce via memory_reinforce_many |

# Feedback flywheel (TAP-1632)

Closes the loop on search misses and explicit agent ratings.

| Action | Parameters | Description |
|---|---|---|
| rate | key, rating, session_id, details_json | Score entry via feedback_rate |

**Note.** Knobs: memory.feedback_auto_emit, memory.feedback_min_similarity in .NLT Quality Pipeline.yaml.

# Security (Epic M1)

Safety and integrity checks before trusting recalled content.

| Action | Parameters | Description |
|---|---|---|
| safety_check | key or query | Content safety scan on memory payload |
| verify_integrity | — | Integrity verification across store + audit trail |

# Profiles and lifecycle (Epic M2)

Brain capability profiles gate which bridge tools are callable.

| Action | Parameters | Description |
|---|---|---|
| profile_info | — | Active profile capabilities |
| profile_list | — | Available profiles |
| profile_switch | profile | Switch active brain profile |

# Health and diagnostics

Surfaced in tapps_session_start brain_bridge_health and tapps doctor.

| Action | Parameters | Description |
|---|---|---|
| health | — | Bridge + native brain health rollup |

# Hive and agent teams (Epic M3)

Cross-agent propagation when CLAUDE_CODE_EXPERIMENTAL_AGENT_TEAMS is enabled.

| Action | Parameters | Description |
|---|---|---|
| hive_status | — | Hive namespaces and agent registry |
| hive_search | query | Search hive-shared entries |
| hive_propagate | key, targets | Propagate entry to agent namespaces |
| agent_register | agent_id, profile, skills | Register agent in hive |

# Pipeline lifecycle (internal)

Phase-2 lifecycle hooks — not daily driver actions on the public catalog.

| Action | Parameters | Description |
|---|---|---|
| session_start_capture | — | Called from tapps_session_start capture path |
| session_end_consolidate | — | Called from tapps_session_end consolidate path |

**Note.** Invoked via session_start_helpers — not advertised as primary agent surface.

## Part V — Graph and hive

PART V

# Knowledge graph and hive

Graph tools (TAP-1630) walk relations and explain paths between entities. Hive tools (Epic M3) propagate entries across agent namespaces when Agent Teams is enabled. Both require HTTP bridge mode for full functionality.

| Surface | Primary actions | Brain tool mapping |
|---|---|---|
| Graph | related, relations, neighbors, explain_connection | memory_find_related, brain_get_neighbors |
| Hive | hive_status, hive_search, hive_propagate, agent_register | Hive namespace registry |
| Federation | federate_* six actions | Hub publish/subscribe/sync |

> “Ralph coordinator cold-start now calls memory_recall_many (≤5 queries) instead of three singleton recalls — then rates relied entries via feedback_rate at debrief.”
>
> **— Ralph coordinator flow (TAP-1887)**

Part VI — Operations

PART VI

# Security, ops, and cross-volume links

Distinguish **dev-time** NLT Quality Pipeline quality pipeline memory from **runtime** NLT Memory consumed by agents during portfolio work. Vol 10 covers maintainer ops; this volume covers engineering reference for the recall authority itself.

| Volume | Link |
|---|---|
| Vol 4 | Memory bridge control-flow sequence diagrams |
| Vol 5 | AgentForge BrainPool during workflow invoke |
| Vol 6 | NLT Engine agents with brain_profile declarations |
| Vol 10 | NLT Quality Pipeline ops — dev-time MCP vs NLT Memory naming |

**Pipeline defaults (POC).** Shipped config enables auto_save_quality, track_recurring_quick_check, auto_supersede_architectural, enrich_impact_analysis, and memory_hooks auto_recall / auto_capture. Override in project .NLT Quality Pipeline.yaml.

# Embeddings and storage layout

Project-local SQLite (.NLT Quality Pipeline/memory/memory.db) holds the MCP-side store for in-process dev workflows: WAL journal, FTS5 full-text search, schema v4 migrations, and optional embedding vectors on entries. Production brain deployments use Postgres with the same tier semantics — the bridge translates MCP actions into native brain tool calls.

| Store | Location | Contents |
|---|---|---|
| Project memory.db | .NLT Quality Pipeline/memory/ | Project-local entries + FTS5 |
| Federated hub | ~/.NLT Quality Pipeline/memory/federated.db | Cross-project shared entries |
| Brain Postgres | NLT Memory service | Authoritative runtime store + graph |
| Audit JSONL | Per-project audit path | Write-through audit trail for recalls |

Search composes BM25 over FTS with optional vector reranking when embeddings are present on entries. Composite ranked search blends relevance, confidence, recency, and access frequency — agents see stable ordering for the same query.

# Operator runbook

Day-one checks before trusting recall in a new checkout. Stub builds omit live package paths and shrink action tables.

| Check | Command / signal | Pass criteria |
|---|---|---|
| Bridge reach | tapps_session_start | brain_bridge_health.ok = true |
| Native health | tapps doctor --quick | Brain row green |
| Profile fit | profile_info | Required actions in capability set |
| Citation verify | report-studio verify | NLT Quality Pipeline paths resolve |

**Build snapshot.** This PDF catalogues **44 actions** in **15 families** from /home/wtthornton/code/NLT Quality Pipeline. Regenerate after MEMORY_REFERENCE.md or server_memory_tools changes.

# Colophon

Generated by reports.nlt_memory_architecture in ReportLab. Structured data from TAPPS_MCP_ROOT and docs/MEMORY_REFERENCE.md.

**NLT Labs · New Logic Tech**
NLT Memory Deep Dive
Issued 2026-06-11 · 44 actions · 15 families