**ADR‑0003 — Session‑per‑Child Model & Sticky Routing (Mcp‑Session‑Id)**

**Document:** docs/adr/0003-session-per-child-sticky-routing.docx  
**Status:** **Accepted**  
**Date:** 2025‑09‑27  
**Project:** MCPX‑KendoBridge  
**Deciders:** DoSE (Accountable), SRE Lead, Dev Lead, Security Lead  
**Consulted:** T‑Arch, DBA, QA Lead, DocFactory  
**Tags:** sessioning, affinity, SSE, child‑process, ingress, readiness, drain

**Guardrails (non‑negotiable):** GitHub‑first SDLC; four environments **Alpha → Beta → RTM (validates on Prod DB read‑only) → Prod**; **Add‑only** schema; **Stored‑procedure‑only** DB access; **No‑Hard‑Coding** of dynamic values (child cmd/args/cwd, timeouts, heartbeat cadence, Origin allow‑list, feature flags) — all read from SQL via sp\_Config\_\*, sp\_Feature\_IsEnabled, sp\_Lookup\_Get; secrets live **only** in GitHub Environments.

**1) Context**

The proxy must bridge Remote MCP clients to the **KendoReact MCP** child process (@progress/kendo-react-mcp) over **STDIO** while supporting **streaming** replies (SSE). We require **isolation**, **predictable performance**, and **observability**. We therefore bind **one child process per session** and keep all requests for that session on the **same replica** to preserve process‑local state and streaming characteristics. This ADR defines the **session key**, affinity mechanism, lifecycle, limits, and failure semantics. It aligns with the **OpenAPI 3.1** surface, NFR budgets (e.g., **SSE TTFB p95 ≤ 200 ms**), and our runbooks.

**2) Decision**

* A **session** is identified by HTTP header **Mcp‑Session‑Id**.
* **On first POST /mcp** without the header, the server **creates** a session, **spawns one child process**, and **returns** the server‑assigned Mcp‑Session‑Id in the response headers.
* All **subsequent requests** for that session **must include** Mcp‑Session‑Id and are **affinitized** to the same replica (**sticky routing**) so they reach the same child process.
* **GET /mcp** opens an SSE **notification** stream **for a session** and **requires** the header.
* The session key is **routing metadata only** (not authentication). Auth (if any) is handled by platform bearer at the gateway.
* Child lifecycle: **created on demand**, **drained on shutdown**, **terminated** when session ends or exceeds idle TTL.
* Failure cases (spawn failure, child unavailable) return canonical error envelopes (spawn\_failed, bad\_gateway\_child\_unavailable) and are observable.  
  These rules are contractually reflected in **OpenAPI 3.1**, FR/NFR, Error Catalog, and runbooks.

**3) Options Considered**

| **Option** | **Pros** | **Cons** |
| --- | --- | --- |
| **Session‑per‑child with header‑based stickiness (chosen)** | Strong isolation; simple mental model; natural for STDIO; easy rollback/drain; good observability | Requires ingress **sticky routing**; more processes under concurrency; needs cleanup/TTL |
| Shared child pool | Fewer processes; possibly lower memory | Harder isolation; cross‑request state bleed; tricky back‑pressure; debugging/rollbacks harder |
| Global singleton child | Simplest to reason | Becomes bottleneck; SPOF; does not scale with sessions |
| WebSocket per client | True bidi; implicit stickiness | Heavier operational surface; proxy compatibility; not required for request→server‑stream pattern |

**4) Rationale**

* **Isolation & predictability.** One child per session avoids **state bleed** and makes failures **session‑scoped**.
* **Operational clarity.** The session header enables **hash‑by‑header** at ingress; runbooks can **drain** streams safely by flipping readiness.
* **Observability.** We can measure session\_count, child\_up, child\_restart\_count, **SSE TTFB**, **heartbeat gaps**, and per‑session lifetimes without logging payload bodies.
* **Compliance fit.** Session behavior is **DB‑configurable** (timeouts, cadence) and documented in /config/effective (non‑secret). **No secrets are stored in DB**.

**5) Implications & Constraints**

**5.1 Ingress & affinity (MUST)**

* Ingress **must not buffer** text/event-stream.
* Sticky routing **must hash on Mcp‑Session‑Id** to keep the session on one replica.
* Read/idle timeouts must allow heartbeats at Network:SseKeepAliveSeconds.  
  (Examples in §10 and runbooks.)

**5.2 Lifecycle & limits**

* **Creation:** First POST /mcp spawns child and issues header.
* **Use:** Subsequent POST /mcp / GET /mcp include Mcp‑Session‑Id.
* **Idle TTL:** Session terminates after a configurable idle period (DB‑sourced key; default TBD).
* **Max sessions/replica:** Target **≥ 200** concurrent sessions/replica (CPU‑bound before memory). Scale‑out if exceeded.
* **Shutdown/drain:** Readiness flips false → SSE streams drain → child terminates → pod exits (see runbooks).

**5.3 Error handling**

* Missing header on GET /mcp → **400 missing\_session\_id**.
* Disallowed origin → **403 origin\_forbidden**.
* Legacy endpoints when disabled → **403 feature\_disabled**.
* Child spawn failure → **500 spawn\_failed**.
* Child crashed/unavailable → **502 bad\_gateway\_child\_unavailable**.  
  (Canonical envelopes per Error Catalog & OpenAPI.)

**6) API & Contract Summary (authoritative)**

* **Headers**
  + **In:** Mcp‑Session‑Id (required for GET /mcp, optional for first POST /mcp).
  + **Out:** Mcp‑Session‑Id (echoed/issued), MCP‑Protocol‑Version (optional future).
* **Endpoints**
  + POST /mcp → JSON (default) or **SSE** (if Accept: text/event-stream).
  + GET /mcp → **SSE** notifications for the session (header required).
  + Legacy (/messages, /sse) **flag‑gated** (EnableLegacyHttpSse).
* **Error envelope**: { code, message, requestId? } on non‑2xx (everywhere).  
  See api/openapi/mcp-proxy.yaml.

**7) Data & Configuration (DB‑sourced, non‑secret)**

* Mcp:ChildCommand, Mcp:ChildArgs, Mcp:ChildCwd (spawn)
* Network:SseKeepAliveSeconds (heartbeat cadence)
* Network:RequestTimeoutSeconds (server timeout)
* Security:AllowedOrigins (Origin allow‑list)
* EnableLegacyHttpSse (feature flag)  
  All values obtained via **SPs** (sp\_Config\_GetValue, sp\_Config\_GetAll, sp\_Feature\_IsEnabled) with **EXECUTE‑only** app grants; **no table CRUD**; **secrets never in DB**.

**8) Observability & SLO Alignment**

* **SLIs/SLOs** (see NFR/Monitoring):
  + Availability ≥ 99.9% monthly; **restart‑to‑ready ≤ 30 s**.
  + **SSE TTFB p95 ≤ 200 ms**; JSON latency p50 ≤ 300 ms / p95 ≤ 800 ms.
  + session\_count per replica; child\_up, child\_restart\_count; heartbeat gap near configured cadence (±1 s).
* **Dashboards:** segment by mode=json|sse, code, session\_count.
* **Alerts:** sustained TTFB or readiness regression; spikes in origin\_forbidden / feature\_disabled.

**9) Testing & Evidence**

**Contract tests:**

* First POST /mcp returns Mcp‑Session‑Id.
* GET /mcp without header → 400 missing\_session\_id.
* Legacy paths return 403 feature\_disabled by default.  
  **E2E (Gherkin):** 01\_session\_establish, 02\_streamed\_tool\_call, 03\_background\_notification, 04\_origin\_denied.  
  **Perf:** capture **SSE TTFB** distribution and heartbeat cadence.  
  **Evidence Pack:** include OpenAPI lint/diff, perf outputs, /ready & /config/effective snapshots, monitoring shots (post‑deploy + 24‑h). Retain **≥ 1 year**.

**10) Ingress & Affinity Examples (non‑secret)**

**NGINX Ingress (Kubernetes annotations)**

nginx.ingress.kubernetes.io/proxy-read-timeout: "3600"

nginx.ingress.kubernetes.io/proxy-send-timeout: "3600"

nginx.ingress.kubernetes.io/proxy-buffering: "off" # critical for SSE

nginx.ingress.kubernetes.io/upstream-hash-by: "$http\_mcp\_session\_id"

nginx.ingress.kubernetes.io/configuration-snippet: |

proxy\_set\_header Connection "";

chunked\_transfer\_encoding off;

**Envoy (stream idle timeout)**

http\_connection\_manager:

stream\_idle\_timeout: 0s

**Note:** Clients should reuse Mcp‑Session‑Id for the life of the session; the server **echoes** the header on successful responses. See Deploy/Scale‑out runbooks for SSE drain and PDB guidance.

**11) Security & Compliance**

* Mcp‑Session‑Id is **not** an auth secret; treat it as routing metadata.
* **Origin allow‑list** enforced from **DB** config (Security:AllowedOrigins), returning 403 origin\_forbidden when violated.
* **No‑Hard‑Coding:** All dynamic values are **DB‑sourced via SPs**; code contains **no literals** for these behaviors.
* **SP‑only:** app has **EXECUTE‑only** grants; **no table rights**; schema is **add‑only**.
* **Secrets policy:** SQL connection strings and **Telerik license** exist **only** in **GitHub Environments**; never in DB/code/logs/evidence.

**12) Backout & Failure Modes**

* **Ingress misconfig (buffering on):** TTFB and heartbeat alerts fire → fix ingress → if needed, **fallback to JSON** on POST /mcp temporarily (no streaming) → verify → re‑enable SSE.
* **Child instability:** observe child\_restart\_count spike → **rollback** to LKG; check DB‑sourced Mcp:Child\* values; verify spawn probe.
* **Affinity failure:** sessions hop between replicas → verify hash‑by‑header; add canary; re‑test probes.  
  All procedures are in Deploy/Rollback/Incident/Scale‑out runbooks with Evidence capture.

**13) Related & Derived Artifacts**

* **ADR‑0001:** Transport choice (Streamable‑HTTP + SSE).
* **ADR‑0002:** Legacy endpoints flag (EnableLegacyHttpSse).
* **OpenAPI 3.1:** api/openapi/mcp-proxy.yaml (headers, SSE examples).
* **NFR & Monitoring:** budgets, SLIs, alerts; RTM parity on **Prod DB (RO)**.
* **Error Catalog:** canonical codes (missing\_session\_id, origin\_forbidden, feature\_disabled, spawn\_failed, bad\_gateway\_child\_unavailable).
* **Runbooks:** deploy / rollback / incident / scale\_out / rotate\_telerik\_license.

**14) Appendices**

**A) Client Header Flow (illustrative)**

POST /mcp # (no Mcp-Session-Id)

← 200 OK

Mcp-Session-Id: s-2f9a...

POST /mcp

Mcp-Session-Id: s-2f9a... # request bound to child

← 200 OK

Mcp-Session-Id: s-2f9a...

GET /mcp

Mcp-Session-Id: s-2f9a... # open SSE notifications for session

← 200 OK (text/event-stream)

**B) Metrics (suggested names)**

* session\_count{pod=…}
* child\_up{pod=…} / child\_restart\_count{pod=…}
* sse\_ttfb\_ms\_bucket{path="/mcp"}
* sse\_heartbeat\_gap\_ms\_bucket{path="/mcp"}
* config\_fetch\_duration\_ms\_bucket{sp="sp\_Config\_\*"}

**Decision record maintained by DocFactory. Changes to sessioning or affinity require synchronized updates to OpenAPI, NFR/Monitoring, runbooks, tests, and this ADR.**