**MCPX‑KendoBridge — Deploy Runbook (Alpha → Beta → RTM → Prod)**

**Document:** runbooks/deploy.docx  
**Version:** 2.0.0 (Kendo Migration)  
**Last Updated:** 2025‑09‑27  
**Owner:** SRE Lead (Responsible) — DoSE (Accountable) — DocFactory (Author)  
**Applies to:** API (.NET 8, Streamable‑HTTP + SSE), Admin Portal (KendoReact Fluent v12, read‑only)

**Purpose.** Provide a **repeatable, auditable** deployment procedure for MCPX‑KendoBridge across **Alpha → Beta → RTM → Prod**, ensuring **SSE pass‑through**, **sticky sessioning** by Mcp‑Session‑Id, **RTM on Prod DB (read‑only)**, and **Kendo license safety** (build‑time only). This runbook aligns with Technijian’s GitHub‑first SDLC, required CI/CD gates, and Evidence Pack retention (≥ 1 year).

**DB & Secrets Compliance (always in effect):**  
**Add‑only** schema; **Stored‑procedure‑only** DAL; **No‑Hard‑Coding** of dynamic values (child cmd/args/cwd, timeouts, heartbeat cadence, allowed origins, feature flags). All such values are **DB‑sourced** via sp\_Config\_\*, sp\_Feature\_IsEnabled, sp\_Lookup\_Get. **Secrets** (SQL connection strings, **Telerik license**) are **never** stored in code/DB/logs; configure them **only** in **GitHub Environments**.

**1) Environments & Endpoints**

| **Env** | **API base** | **UI base** | **DB** | **Notes** |
| --- | --- | --- | --- | --- |
| **Alpha** | https://alpha.example.com/api | https://alpha.example.com/ | Alpha | Functional bring‑up |
| **Beta** | https://beta.example.com/api | https://beta.example.com/ | Beta | Hardening + perf smoke |
| **RTM** | https://rtm.example.com/api | https://rtm.example.com/ | **Prod DB (read‑only)** | Parity gate |
| **Prod** | https://prod.example.com/api | https://prod.example.com/ | Prod | Canary → rollout; 24‑h checks |

**Ingress requirement:** **Do not buffer** text/event-stream; honor idle/read timeouts to allow heartbeats at Network:SseKeepAliveSeconds (DB‑sourced). Sticky routing by **Mcp‑Session‑Id** header is required for active sessions.

**2) Roles & RACI (Deployment)**

| **Activity** | **A** | **R** | **C** | **I** |
| --- | --- | --- | --- | --- |
| Release approval & readiness | DoSE | SRE Lead | Dev Lead, QA Lead, SecLead | All |
| GitHub Actions deploy | DoSE | SRE | Dev Lead | QA |
| SSE/TTFB validation | DoSE | SRE | QA | Dev |
| Evidence Pack attachment | DoSE | SRE/CI | QA | All |
| 24‑h post‑release checks | DoSE | SRE + QA | Dev Lead | All |

**3) Change Types**

* **Standard release** (preferred): image from merge‑queue‑green main; promotes Alpha → Beta → RTM → Prod.
* **Hotfix** (exceptional): same flow; add **post‑mortem** to Evidence Pack.
* **Rollback**: see runbooks/rollback.docx (graceful SSE drain; image flip).

**4) Pre‑Deploy Checklist (must be all green)**

**CI gates** (PR or main): Build/Tests, **OpenAPI lint/diff**, **CodeQL** (C#/JS), **Dependency Review** (no **High**), **Secret Scanning** (no leaks), **SBOM** artifact.  
**Artifacts present:** test results, SBOM, CodeQL SARIF, OpenAPI files, UI build (if exists).  
**DB compliance:** migrations are **add‑only**; SP signatures unchanged; app **EXECUTE‑only** on SPs.  
**SSE ingress:** verify **no buffering** for text/event-stream.  
**Sticky sessions:** routing keyed by Mcp‑Session‑Id.  
**Feature flags:** EnableLegacyHttpSse remains **false** unless a specific interop requires it (document if toggled).  
**Kendo license:** present **only** in GitHub Environments for UI build; not in images/files.  
**Approvals:** Environment approvals configured (Alpha/Beta/RTM/Prod).  
**Image:** tag or digest known (immutable digest preferred).  
**Evidence plan:** release Evidence Pack structure ready.

**5) Deployment via GitHub Actions**

**Workflow:** /.github/workflows/deploy.yml  
**Inputs:** image (tag/digest), environment (alpha|beta|rtm|prod)

**5.1 Alpha**

1. **Dispatch:** Provide image + environment=alpha.
2. **Wait for job:** Deploy Alpha.
3. **Validate (API):**
4. curl -fsS https://alpha.example.com/api/ready | jq .
5. curl -fsS https://alpha.example.com/api/healthz | jq .
6. curl -fsS https://alpha.example.com/api/config/effective | jq .
7. **Validate (SSE TTFB):**
8. # First-byte time (TTFB) in ms
9. curl -N -H 'Accept: text/event-stream' \
10. -H 'Mcp-Session-Id: alpha-check' \
11. -H 'Content-Type: application/json' \
12. -w '\nTTFB(ms)=%{time\_starttransfer}\n' \
13. -d '{"jsonrpc":"2.0","id":"1","method":"ping","params":{"stream":true}}' \
14. https://alpha.example.com/api/mcp
15. **UI sanity (if present):** load / → Dashboard tiles render (uptime, sessionCount, childProcesses).
16. **Evidence:** upload Alpha snapshots (/ready, /healthz, /config/effective) to the release draft under evidence/05\_ops.
17. **Decision:** proceed to **Beta** if green.

**5.2 Beta**

1. **Dispatch:** environment=beta.
2. **Perf smoke:** run script (latency + streaming TTFB); expect budgets met (JSON p50≤300 ms/p95≤800 ms; **SSE TTFB p95 ≤ 200 ms**).
3. **Policy check:** send a request with disallowed Origin; expect 403 origin\_forbidden (envelope).
4. **Evidence:** upload perf results to evidence/06\_perf.

**5.3 RTM (validates on Prod DB read‑only)**

1. **Dispatch:** environment=rtm.
2. **Parity:** capture /config/effective and compare to expected Prod values (non‑secret). **Block** if drift detected.
3. **Contract tests:** run OpenAPI contract tests; **no writes** to DB.
4. **Evidence:** attach parity diff and contract test outputs.

**5.4 Prod (canary → full rollout)**

1. **Dispatch:** environment=prod.
2. **Canary:** deploy **10%**; observe for 15–30 min (latency, error rate, **TTFB**, readiness).
3. **Rollout:** to 100% if stable.
4. **24‑hour checks:** schedule and later attach monitoring snapshots (availability ≥99.9%, error rate <1%, restart‑to‑ready ≤30 s, **TTFB p95 ≤ 200 ms**).
5. **Finalize Evidence Pack** and publish release.

**6) Manual Deployment (fallback)**

If GitHub Actions is unavailable:

1. **Container update** via platform tooling (Kubernetes/VM).
2. **Readiness gate:** ensure /ready flips to fail before terminating old pods (drain).
3. **SSE drain:** wait for active SSE streams to complete before killing the pod (graceful).
4. **Sticky routing:** maintain Mcp‑Session‑Id affinity during rolling update.
5. **Run smoke tests** as in §5.

**7) SSE Pass‑Through & Sticky Sessions — Reference**

**NGINX Ingress (example annotations):**

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

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

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

location /api/mcp {

proxy\_set\_header Connection "";

chunked\_transfer\_encoding off;

proxy\_buffering off; # critical for SSE

}

**Envoy (snippet):**

http\_filters:

- name: envoy.filters.http.router

typed\_config: {}

# For SSE

stream\_idle\_timeout: 0s

**Sticky routing (concept):** hash or cookie affinity on Mcp-Session-Id so that requests for a session land on the **same replica** while active. Consult your ingress/load balancer docs to hash by a header.

**8) Kendo License Handling (build‑time only)**

* The Admin Portal uses KendoReact. CI detects @progress/kendo-\* dependencies and **requires** a **TELERIK\_LICENSE** (or TELERIK\_LICENSE\_PATH) **at build time**.
* CI writes the license to a temporary file ($RUNNER\_TEMP), sets TELERIK\_LICENSE\_PATH, builds the UI, and **does not** persist the license in logs/images/artifacts.
* **Never** commit or store the license in DB or repo. For rotation, follow runbooks/rotate\_telerik\_license.docx.

**9) Validation Matrix (per environment)**

| **Check** | **Alpha** | **Beta** | **RTM** | **Prod** |
| --- | --- | --- | --- | --- |
| /ready & /healthz green | ✓ | ✓ | ✓ | ✓ |
| /config/effective snapshot (non‑secret) | ✓ | ✓ | ✓ | ✓ |
| Perf smoke (JSON p50/p95; **TTFB**) | — | ✓ | optional | ✓ |
| 403 origin\_forbidden on bad Origin | ✓ | ✓ | ✓ | ✓ |
| Legacy endpoints off (403 feature\_disabled) | ✓ | ✓ | ✓ | ✓ |
| RTM parity against Prod values | — | — | ✓ | — |
| Canary (5–10%), then rollout | — | — | — | ✓ |
| 24‑hour post‑release checks & evidence | — | — | — | ✓ |

**10) Evidence Pack Hooks (what to attach)**

* **Contracts:** OpenAPI, lint/diff, Error Catalog.
* **Tests:** TRX/coverage, E2E Gherkin outputs, axe report, contract tests.
* **Security:** CodeQL SARIF, Dependency Review, Secret‑scan summary.
* **Supply chain:** SBOM + image digest.
* **Ops:** /ready, /healthz, /config/effective per env; **RTM parity** diff.
* **Perf:** latency & **TTFB** results; heartbeat cadence.
* **Monitoring:** post‑deploy and **24‑h** screenshots; approval snapshots. Retain ≥ 1 year.

**11) Troubleshooting (quick checks)**

* **SSE stalls / missing heartbeats:**
  + Verify ingress **proxy\_buffering off** (NGINX) or equivalent; check read/idle timeouts.
  + Confirm Network:SseKeepAliveSeconds via /config/effective.
* **Sessions flap / tool calls fail:**
  + Check sticky routing by Mcp‑Session‑Id.
  + Inspect child\_up / child\_restart\_count metrics.
* **403 on allowed origin:**
  + Validate Security:AllowedOrigins in DB → /config/effective; beware scheme/host/port mismatches.
* **Legacy endpoint unexpectedly on/off:**
  + Check EnableLegacyHttpSse flag; default is **false**.
* **Readiness fails after rollout:**
  + DB SP reachability or child spawn probe failing; check logs (no payload bodies).
* **UI build fails (Kendo):**
  + Ensure TELERIK\_LICENSE or \_PATH set in the target GitHub Environment; verify CI step.

**12) Safety & Backout**

If any **P1** condition occurs (availability <99% 10‑min window, /ready failing >2 min, **TTFB p95 > 200 ms** sustained), **pause rollout** and follow runbooks/incident.docx; if needed, execute runbooks/rollback.docx (graceful SSE drain, image flip). Post‑mortem and Evidence updates required.

**13) Appendices**

**A) Curl snippets**

* **JSON mode**
* curl -fsS -H 'Content-Type: application/json' \
* -d '{"jsonrpc":"2.0","id":"1","method":"ping","params":{}}' \
* https://<env>.example.com/api/mcp | jq .
* **SSE mode**
* curl -N -H 'Accept: text/event-stream' -H 'Content-Type: application/json' \
* -H 'Mcp-Session-Id: test' \
* -d '{"jsonrpc":"2.0","id":"1","method":"ping","params":{"stream":true}}' \
* https://<env>.example.com/api/mcp

**B) Required GitHub Environment Secrets (non‑exhaustive)**

| **Env** | **Secret** | **Purpose** |
| --- | --- | --- |
| alpha/beta/prod | SQL\_CONNECTION\_STRING | API database connection |
| rtm | SQL\_CONNECTION\_STRING\_PROD\_RO | **Prod DB read‑only** for RTM parity |
| any with UI | TELERIK\_LICENSE or TELERIK\_LICENSE\_PATH | Kendo UI build‑time licensing |

**C) References**

* **CI/CD Plan** (docs/10\_ci\_cd.docx) — workflows, gates, promotion.
* **Compliance** (docs/13\_compliance.docx) — secrets policy, CSP/egress, DB rules.
* **Monitoring** (docs/11\_monitoring.docx) — SLOs, alerts, **TTFB** dashboards.
* **Error Catalog** (docs/error\_catalog.docx) — canonical error codes.
* **OpenAPI** (api/openapi/mcp-proxy.yaml) — transport, headers, examples.

**Footer (optional for Word header/footer):**  
*MCPX‑KendoBridge • Deploy Runbook • v2.0.0 • 2025‑09‑27 • Confidential — Technijian Internal*