Cloud Migration 2026

ALAI cloud migration master plan: 6-phase transition from ANVIL-only to cloud-hosted control plane

Master Plan — Cloud Migration

$(cat /tmp/bookstack-page-1-master-plan.html | jq -Rs .)

Phase 1 — Bitwarden Cloud Migration

Phase 1 — Bitwarden Cloud Migration

Timeline: Days 1-3
Goal: Eliminate Vaultwarden SPOF as the very first step. Every subsequent phase depends on secrets being available globally, not just when the Azure VM is alive.
MC Task: #8494
Proveo Owner: Angie Jones
Status: PREVIEW — Parisa writing detailed runbook in parallel

Why First

Phase 2 onwards deploys to Azure Container Apps. Those containers need secrets at startup (Anthropic API key, Postgres connection string, Azure SP). If Vaultwarden is down, all containers fail to start. Fix the foundation before building on it.

Deliverables

- Export all current Vaultwarden items to encrypted JSON
- Import to Bitwarden cloud Teams ($4/user/month — 1 seat = $4/month total)
- Update `alai-cli` bootstrap step to use `bw login` against `cloud.bitwarden.com`
- Update all agent bootstrap scripts to use cloud BW endpoint
- Delete the BW CLI config pointing to `vault.basicconsulting.no`

Rollback Plan

Vaultwarden self-hosted remains running in parallel until Phase 6. If Bitwarden cloud import fails, fall back to self-hosted immediately. Keep vault export as encrypted offline backup in `~/system/backups/`.

Proveo Validation Criteria

Test Owner: Angie Jones (Proveo)

1. Fresh `bw login alembasic@gmail.com` on a machine with NO `vault.basicconsulting.no` access returns all expected items (GitHub token, Azure SP, Anthropic key, SSH key)
2. `alai login` (once built in Phase 4) succeeds using cloud BW credentials
3. Vaultwarden VM can be stopped for 1 hour with no agent failures on ANVIL

Cost

Bitwarden cloud Teams: $4/user/month × 1 user = $4/month
vs Vaultwarden HA (2 VMs + Load Balancer): ~$88/month

Detailed Runbook

Parisa Tabriz (Securion) is writing the full step-by-step runbook in parallel. Once complete, it will be referenced here:
`~/system/architecture/phase-1-bitwarden-runbook.md` (pending)

Phase 2 — MC + HiveMind API

Phase 2 — MC + HiveMind API

Timeline: Weeks 1-2
Goal: Mission Control and HiveMind leave ANVIL and become cloud-hosted APIs. This is the biggest architectural change — SQLite becomes Postgres, local scripts become REST calls.
MC Task: #8495
Proveo Owner: Angie Jones
Status: PREVIEW — Kelsey working in parallel

Why Second

MC and HiveMind are the nervous system. Once they are cloud-hosted, every other phase can run from any machine without touching ANVIL.

Deliverables

- **mc-api.js:** Express-based REST API wrapping current `mc.js` logic
  - `GET /tasks`, `POST /tasks`, `PATCH /tasks/:id`, `GET /stats`
  - Postgres driver (pg) replacing SQLite
  - Schema migration: 8378 tasks, 127 open — pg-migrate from SQLite dump
- **hivemind-api.js:** REST + optional WebSocket for pub/sub
  - Postgres backend (hivemind schema)
- Docker images for both, pushed to Azure Container Registry
- **Azure Container Apps:** deploy mc-api and hivemind-api
  - Consumption plan (serverless, scale-to-zero when no traffic)
  - Min replicas: 1 (so cold start is 2-4s max, not 30s+)
  - Memory: 0.5GB each, vCPU: 0.25 each
- **Azure Database for Postgres Flexible Server:** Burstable B1ms
  - Region: swedencentral
  - `mission_control` DB + `hivemind` DB on same instance
  - Automated backups (7-day retention, included in cost)
- Update `mc.js` client wrapper: detect `ALAI_MC_URL` env var, proxy to API if set
  - Backward compatible: if no `ALAI_MC_URL`, still uses local SQLite (ANVIL stays working)

Cost Estimate

```
Container Apps (2 apps, ~5h/day active, consumption plan):
  ~$1.50/month per app = $3/month total
  (Free grant: 180,000 vCPU-s/month covers most light usage)

Azure Postgres B1ms: ~$22-24/month (swedencentral, Flexible Server)
Azure Container Registry Basic: $5/month

Total Phase 2 additions: ~$30-32/month
```

Rollback Plan

`mc.js` still reads local SQLite if `ALAI_MC_URL` is not set. If Postgres or Container Apps fail, unset `ALAI_MC_URL` on ANVIL and operations continue locally. SQLite is kept in parallel for 30 days post-migration before decommission.

Proveo Validation Criteria

Test Owner: Angie Jones (Proveo)

1. From ab-mac (no local SQLite): `alai mc list` returns live tasks
2. From ANVIL: `node ~/system/tools/mc.js list` still works (backward compat)
3. POST to mc-api: task appears in both `mc.js list` AND cloud Postgres within 2s
4. Postgres automated backup: verify restore of 100-row sample matches source
5. Container App scales to zero after 10min idle, cold starts under 5s

Detailed Implementation

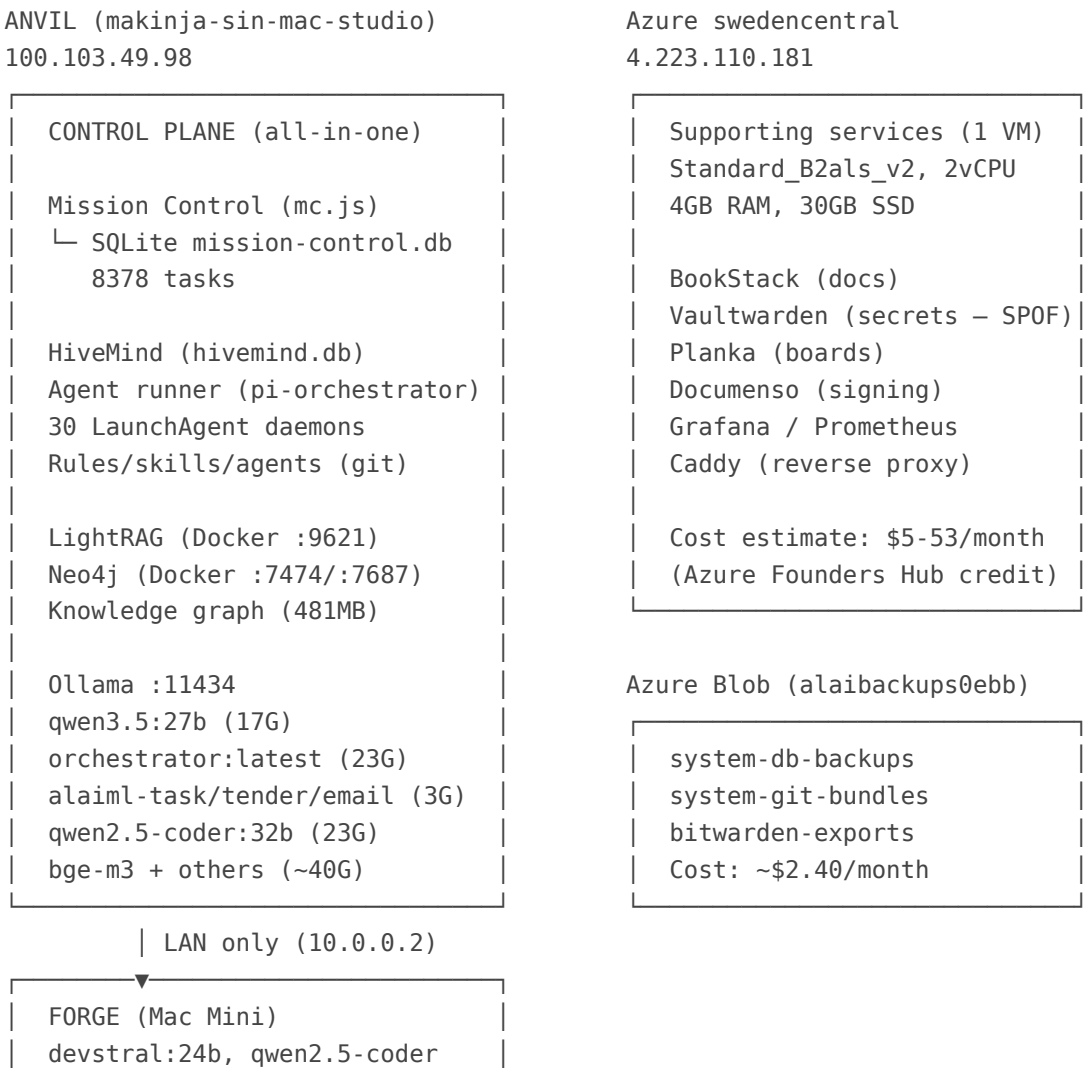
Kelsey Hightower (FlowForge) is implementing Azure Container Apps + Postgres in parallel. Full runbook will be linked here once ready.

Current State vs Target State

Current State vs Target State

Purpose: Visual comparison of ALAI's architecture today (ANVIL single-point-of-failure) vs the cloud-hosted control plane target state.
Source: `~/system/architecture/cloud-migration-master-plan.md`

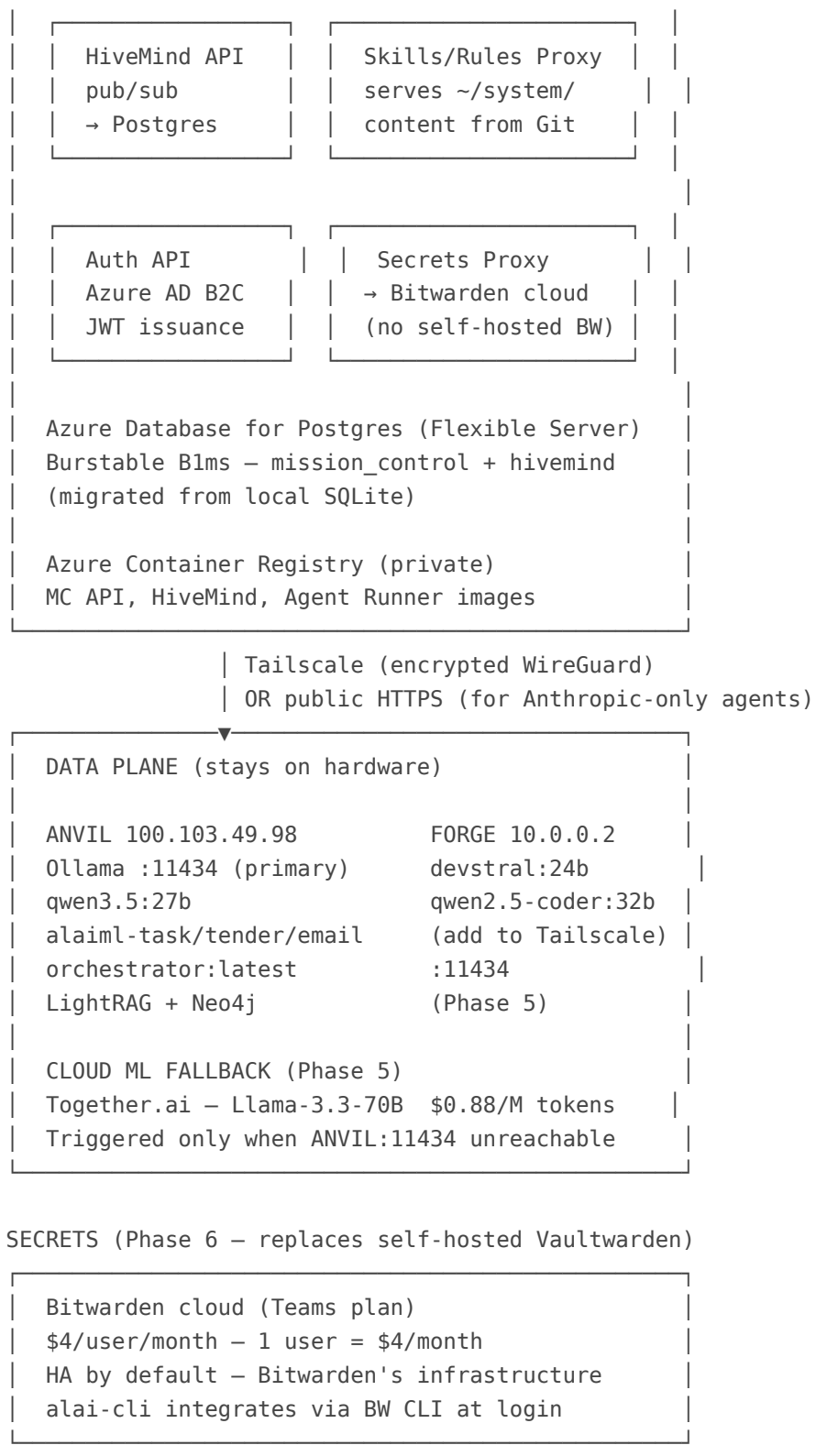
TODAY — SINGLE SPOF ARCHITECTURE



```
  │  NOT on Tailscale — LAN only    │
  └─────────────────────────────────┘

  Tailscale mesh: 4 nodes
    makinja-sin-mac-studio  100.103.49.98
    ab-mac                  100.118.37.71
    basicass-mac-mini       100.104.164.86
    iphone181               100.93.161.73

  NOTE: ANVIL Ollama :11434 NOT reachable from ab-mac (port timeout verified).
  NOTE: 306 files in ~/system/ hardcode localhost:11434 — zero portability today.

SPOF inventory (4 critical):
  [1] ANVIL dead       → mc.js, HiveMind, agents, LightRAG, Ollama ALL stop
  [2] FORGE dead       → devstral/coder workload stops (Anthropic can substitute)
  [3] Azure VM dead    → Vaultwarden down, secrets inaccessible, agents cannot bootstrap
  [4] Local network    → FORGE permanently isolated (LAN-only, no Tailscale)
```

TARGET — CLOUD-HOSTED CONTROL PLANE + THIN CLIENT

```
  CLIENT (any OS — new laptop, travel machine, etc.)
  ┌──────────────────────────────────────────────────────┐
  │  alai-cli (single installable package)               │
  │  brew install alai  |  npm install -g @alai/cli      │
  │  winget install alai  |  apt install alai-cli        │
  │                                                      │
  │  alai login     → OAuth2 PKCE → Azure AD B2C         │
  │  alai start     → connects to cloud APIs             │
  │  alai mc list   → proxies to MC API                  │
  │  alai agent run → dispatches to agent runner         │
  │                                                      │
  │  Claude Code CLI (installed separately)              │
  │  ~/.claude/ cloned from git on login                 │
  └──────────────────────────────────────────────────────┘
                  | HTTPS (Azure Front Door or direct)
                  | Auth: Azure AD B2C JWT
  ┌───────────────▼──────────────────────────────────────┐
  │  CLOUD CONTROL PLANE (Azure Container Apps)          │
  │  Region: swedencentral (existing subscription)       │
  │                                                      │
  │  ┌──────────────────┐  ┌──────────────────────────┐  │
  │  │  MC API          │  │  Agent Runner API        │  │
  │  │  REST + WebSocket│  │  POST /run               │  │
  │  │  → Postgres      │  │  → dispatches agents     │  │
  │  └──────────────────┘  └──────────────────────────┘  │
  │                                                      │
```



Key Differences

Component	Current State (ANVIL SPOF)	Target State (Cloud Control Plane)

Mission Control	SQLite on ANVIL disk	Postgres + MC API (Azure Container Apps)
HiveMind	SQLite on ANVIL disk	Postgres + HiveMind API (Azure Container Apps)
Agent Runner	pi-orchestrator on ANVIL only	Cloud agent-runner (Anthropic-powered agents), ANVIL for fine-tuned models
Secrets	Vaultwarden on single Azure VM	Bitwarden cloud ($4/month, HA by default)
Client Bootstrap	Manual setup, ANVIL-dependent	`brew install alai && alai login` — under 10 minutes, any OS
Ollama	ANVIL only, FORGE LAN-isolated	ANVIL + FORGE (Tailscale) + Together.ai cloud fallback
Cost	$27-106/month (mostly hidden by Azure credit)	$108-165/month (transparent, no hidden dependencies)
ANVIL Offline Impact	Total system outage	Cloud services continue, fine-tuned models pause gracefully

SPOF Elimination

4 SPOFs removed:

1. **ANVIL death** — control plane (MC, HiveMind, agent runner) migrates to cloud. ANVIL offline = Ollama workloads pause, everything else continues.
2. **Vaultwarden VM death** — secrets migrate to Bitwarden cloud (HA by default). No more single-VM secret dependency.
3. **Network isolation** — FORGE joins Tailscale. Cloud services can reach FORGE for code tasks even when ANVIL is down.
4. **Workstation lock-in** — `alai-cli` works from any machine. No more "John only works from ANVIL."

Credit: ALAI, 2026