Bilko Self-Serve Trial — CIAM Architecture and Auth Pattern (MC #103232)

Bilko Self-Serve Trial — CIAM Architecture & Auth Pattern

MC: #103232 | **Status:** LIVE — Proveo 11/11 PASS | **Last updated:** 2026-06-09 | **Securion verdict:** LAUNCH WITH CONDITIONS

---

1. Overview

A prospect navigates to `app.bilko.cloud` (or `bilko-demo.alai.no`), clicks **"Sign in or create a free account with your email"**, and completes a Microsoft CIAM Email-OTP sign-up. On first login, the backend JIT-provisions an empty Bilko organisation with a **7-day trial** directly on the real production database (`bilko-demo-db`). There is no separate demo build, no invite-only flow, and no org Microsoft account required — any personal email address works.

The deployment target is the standard `bilko-main-deploy` semver-tag trigger. Stage and demo share the same Kotlin/Ktor binary and the same database instance (multi-tenant via RLS). The CIAM tenant (`bilkociam`) is a dedicated Microsoft Entra External ID tenant, completely separate from the Bilko staff Entra tenant.

1.1 Flow diagram

```
Prospect → app.bilko.cloud/login
         → "Sign in with Microsoft" (MSAL redirect)
         → bilkociam.ciamlogin.com [BilkoSignUpSignIn user flow]
         → Email OTP verification (8-digit code, ~6s delivery)
         → Consent pages (2 pages on first login only)
         → Redirect to app.bilko.cloud/auth/callback
```

```
→ MSAL: LOGIN_SUCCESS fires, payload.idToken available
→ POST /auth/entra/session { idToken }   [B1 exchange fix]
→ bilko-api: JWKS RS256 verify → OID lookup → JIT provision
→ Response: Bilko HMAC JWT + org { trialEndsAt }
→ setAuthFromRegistration()              [B1.2 session fix]
→ checkAuth() in-memory JWT fast-path    [B1.3 session fix]
→ /dashboard — empty org, trial active ("Probno: 6 dana preostalo")
```

---

2. CIAM Tenant Configuration

Property	Value
Tenant name	bilkociam
Tenant ID	20bb17de-9be5-4143-a7e5-8c1ddae6a064
Tenant type	CIAM (Entra External ID)
SPA app name	Bilko Web (SPA)
SPA client ID	c2902239-ea63-41bd-8619-6cf096d7d45a
API resource app ID	fe39e0f5-513e-40af-93f0-c3ee624df56c
Authority URL	https://20bb17de-9be5-4143-a7e5-8c1ddae6a064.ciamlogin.com/20bb17de-9be5-4143-a7e5-8c1ddae6a064/v2.0
OIDC issuer	same as authority URL (confirmed via discovery endpoint)

2.1 User flow: BilkoSignUpSignIn

Property	Value
Flow ID	aa86084b-01dc-453f-9e10-679dfefdd824
Type	externalUsersSelfServiceSignUpEventsFlow
Display name	BilkoSignUpSignIn
Identity provider	EmailOtpSignup-OAUTH (Email One Time Passcode)
isSignUpAllowed	true
userTypeToCreate	member (not guest)
Attributes collected	email (auto-filled by OTP verification)
Linked app	c2902239-ea63-41bd-8619-6cf096d7d45a (Bilko Web SPA)

Authority URL note: Unlike Azure AD B2C, Entra External ID CIAM does *not* require a user flow policy name suffix in the authority URL. The BilkoSignUpSignIn flow is applied automatically at the

tenant level when the SPA app is linked to it. The deployed authority URL requires no changes.

2.2 Registered SPA redirect URIs

- https://app.bilko.cloud/auth/callback and https://app.bilko.cloud
- https://app.bilko.company/auth/callback and https://app.bilko.company
- https://app.bilko.io/auth/callback and https://app.bilko.io
- https://bilko-demo.alai.no/auth/callback and https://bilko-demo.alai.no
- https://bilko-web-stage-dh4m46blja-lz.a.run.app/auth/callback and .a.run.app
- http://localhost:3000/auth/callback and http://localhost:3000

2.3 Adding identity providers or attributes

To add social identity providers (Google, Apple) or additional signup attributes (e.g. display name, company name): Microsoft Entra admin centre → External Identities → User flows → BilkoSignUpSignIn → Identity providers / Attributes. No code changes or redeploys are required for attribute-only changes. Adding a social provider requires app registration on the provider side and linking in the CIAM tenant.

---

3. Auth Flow — The Hard-Won Pattern

This section documents three bugs that were discovered and fixed during Proveo E2E validation (MC #103232 WS-V). The fixes are canonical — do not revert them.

B1 — Token exchange (commit 660f410, tag v0.2.45)

Problem: MSAL's `LOGIN_SUCCESS` event fires with an Entra `access_token` (RS256, Microsoft-issued). The original code set this directly as the API Bearer header. The Bilko API validates HMAC256 JWTs only — all calls returned 401.

Fix: After MSAL fires, pass `payload.idToken` (not `payload.accessToken`) to a `POST /auth/entra/session { idToken }` call. The backend verifies the CIAM RS256 idToken via JWKS, looks up or JIT-provisions the user, and returns a Bilko HMAC JWT.

```
// apps/web/lib/msal/msal-provider.tsx — corrected token selection
const idToken = payload.idToken ?? payload.accessToken
if (idToken) { handleEntraLogin(idToken) }

// apps/web/lib/msal/use-entra-auth.ts — exchange call
```

```
const sessionResult = await api.auth.entraSession(idToken)
const bilkoJwt = sessionResult?.tokens?.accessToken
setAccessToken(bilkoJwt)
```

B1.2 — Session persistence via setAuthFromRegistration (commit e1e31c5, tag v0.2.46)

Problem: After the B1 exchange, `checkAuth()` was called to hydrate the store. `checkAuth()` internally calls `POST /auth/refresh` using the httpOnly refresh-token cookie. The CIAM exchange path does not set a cookie — so `/auth/refresh` returned 401, which cleared the Bilko JWT and redirected back to `/login`.

Fix: Replace the `checkAuth()` call in `handleEntraLogin` with `setAuthFromRegistration()`, which hydrates the Zustand auth store directly from the `/auth/entra/session` response body. No cookie round-trip needed.

```
// apps/web/lib/msal/use-entra-auth.ts — hydrate from session response
const { setAuthFromRegistration } = useAuthStore.getState()
setAuthFromRegistration({
  user: sessionResult.user,
  organization: sessionResult.organization,
  tokens: { accessToken: bilkoJwt },
})
// Navigate to /dashboard — Bilko JWT is in-memory Bearer
```

B1.3 — checkAuth in-memory JWT fast-path (commit 30a8c85, tag v0.2.47)

Problem: Even with B1.2, `setAuthFromRegistration()` in `handleEntraLogin` correctly set `isAuthenticated=true`. However, `AuthProvider` mounts on every protected route and calls `checkAuth()`. That call hit `/auth/refresh` (cookie path) → 401 → store reset to unauthenticated → redirect to `/login` on every page navigation.

Fix: Added an in-memory JWT fast-path at the top of `checkAuth()` in `auth-store.ts`. If a Bilko JWT is already in memory (set via the CIAM exchange), `checkAuth()` uses `GET /auth/me` with that Bearer token instead of falling through to the cookie-refresh path.

```
// apps/web/lib/stores/auth-store.ts — in-memory fast-path
checkAuth: async () => {
  const inMemoryToken = getAccessToken()
```

```
  if (inMemoryToken) {
    try {
      const me = await api.auth.me()
      set({ isAuthenticated: true, isLoading: false,
            user: { ...me, name: me.fullName },
            organization: me?.organization ?? null })
      return true
    } catch {
      set({ isAuthenticated: false, isLoading: false, user: null, organization: null })
      return false
    }
  }
  // Original cookie-refresh fallback (unchanged — non-CIAM sessions)
  ...
}
```

ENTRA_EXTERNAL_ID_AUDIENCE — critical build var (fixed in v0.2.47 trigger update)

Problem: The `bilko-main-deploy` Cloud Build trigger had `_ENTRA_EXTERNAL_ID_AUDIENCE` set to `fe39e0f5` (the API resource app ID). This is wrong — the CIAM idToken audience is the *SPA client ID* (`c2902239`), because MSAL requests id_tokens scoped to the requesting app. Every new deploy reverted the Cloud Run env to the wrong value, requiring a manual patch.

Fix: The trigger substitution was updated:

```
# infrastructure/gcp/cloudbuild.yaml — correct value
_ENTRA_EXTERNAL_ID_AUDIENCE: c2902239-ea63-41bd-8619-6cf096d7d45a   # SPA client ID
# NOT: fe39e0f5-513e-40af-93f0-c3ee624df56c  (that is the API resource app — wrong for idToken a
```

This is now stable in the trigger — it will not revert on future deploys.

---

4. Backend JIT Provisioning

4.1 Database migrations (Flyway V66–V69)

Migration	Purpose
V66__entra_rls_and_password_nullable.sql	Makes `password_hash` nullable (Entra-only users have no password). Adds RLS policy on `entra_external_identities` (FORCE + fail-closed). Adds CHECK constraint: issuer must not end with trailing slash.

Migration	Purpose
V67__rbac_permissions_catalog.sql	RBAC permissions catalog seeding.
V68__rbac_user_provisioning.sql	SECURITY DEFINER function `bilko_auth.provision_user_with_org()`: creates org (7-day trial, `trial_starts_at`, `trial_ends_at = now() + 7 days`), creates user (`role='viewer'`, `password_hash=NULL`), inserts `entra_external_identities` row (issuer, OID, user_id). Default: `country='BA'`, `currency='BAM'`.
V69__fix_provision_rls.sql	RLS fix: calls `set_config('app.current_org_id', v_org_id, true)` before the users INSERT so that RLS policies on the users table pass during JIT provisioning.

4.2 JIT provisioning call flow

```
POST /auth/entra/session { idToken }
  → EntraExternalIdService.verifyIdToken()         [RS256, JWKS, issuer+audience+exp]
  → AuthUserRepository.findByEntraIdentity(issuer, oid)  → null (new user)
  → AuthUserRepository.findByEmail(email)                → null (no existing Bilko account)
  → UserProvisioningService.provisionNewUserForEntra()
      → bilko_auth.provision_user_with_org()  [SECURITY DEFINER, SERIALIZABLE]
          → INSERT organizations (trial 7 days)
          → INSERT users (role=viewer, password=null)
          → INSERT entra_external_identities (issuer, oid)
  → jwtService.signAccessToken(userId, email, role='viewer', orgId)
  → Response: { user, organization { trialEndsAt }, tokens { accessToken, refreshToken } }
```

Idempotency: Re-login with the same OID returns the existing user and org; trial end date is not reset. The `entra_external_identities` table has a UNIQUE constraint on `(issuer, subject)`.

4.3 trialEndsAt in /auth/me

The `GET /auth/me` response includes `organization.trialEndsAt` (ISO 8601). The frontend auth store exposes this on the `Organization` interface. The trial expiry is enforced server-side by `TrialGatePlugin` which queries the DB on every gated request — the JWT does not embed expiry.

4.4 RLS isolation

All JIT-provisioned tenants are isolated via PostgreSQL Row Level Security. The `app.current_org_id` session variable is set by `OrgScopePlugin` from the `BilkoPrincipal` (JWT-derived, not from any HTTP header). `orgTransaction()` uses `SET LOCAL` scoped to the transaction — connection pool does not carry state between requests. Cross-tenant isolation is verified by `RlsOrgIsolationV46IntegrationTest`.

---

5. Deploy

Property	Value
Deploy trigger	bilko-main-deploy (europe-north1, project tribal-sign-487920-k0)
Trigger type	semver tag on main: `git tag vX.Y.Z && git push origin vX.Y.Z`
Config	infrastructure/gcp/cloudbuild.yaml
Current live tag	v0.2.47 (commit 30a8c85)
Web revision	bilko-web-demo-00080-tq5
API revision	bilko-api-demo-00155-524
CIAM env vars in trigger	NEXT_PUBLIC_ENTRA_CLIENT_ID, NEXT_PUBLIC_ENTRA_AUTHORITY, NEXT_PUBLIC_ENTRA_SCOPE, ENTRA_EXTERNAL_ID_ISSUER, ENTRA_EXTERNAL_ID_AUDIENCE (= c2902239), ENTRA_EXTERNAL_ID_JWKS_URL

ZAKON PI2: Do not run `cloudbuild.yaml` manually. Use `git tag + git push origin` only. The stage pipeline (`bilko-stage-auto-deploy`) fires on every push to main and is unrelated to the demo deploy.

---

6. Known Follow-Ups

ID	Priority	Description
H1	HIGH — must-fix before scale launch	**Abuse gate (MC #103245):** JIT provisioning has no server-side rate gate on tenant creation. An attacker with many email inboxes can script CIAM sign-ups (each requires a real OTP but automation services exist). Fix: add a platform-level provision rate gate in `UserProvisioningService.provisionNewUserForEntra()` (max N JIT orgs per hour) + CIAM tenant configuration to block disposable email domains.

ID	Priority	Description
B3	MEDIUM	**Migadu email OTP blocking:** Migadu (one.com), used for `@alai.no`, blocks Microsoft Azure CIAM OTP emails. Prospects with Gmail or Outlook receive OTP in ~6 seconds. Alai staff using `@alai.no` addresses cannot sign up. Fix: whitelist `accountprotection.microsoft.com` sender in Migadu SPF settings, or configure a custom CIAM email sender domain.
UX-1	LOW	**Org display name:** JIT-provisioned orgs are named "unknown's Organization" (no display name collected at signup). The user flow only collects email. Fix: add displayName to the BilkoSignUpSignIn attribute collection (Azure config, no code change), or collect it on first post-login screen.
UX-2	LOW	**Default country/currency:** JIT-provisioned org defaults to `country='BA'`, `currency='BAM'`. Prospects outside Bosnia must update via Settings. A country selection step at signup would improve the onboarding experience (follow-on, not a blocker).
M1	MEDIUM	**INGRESS_TRAFFIC_ALL (MC #99924):** Direct `*.run.app` access bypasses GCLB, which degrades IP-based rate limiting to per-GFE-region keying. Pre-existing risk, not introduced by CIAM. Fix: lock ingress to internal-only when load balancer is provisioned.
M3	MEDIUM	**No alert on rapid tenant creation:** Add a GCP Cloud Monitoring alert triggering when more than N organisations are JIT-provisioned per hour.

7. Validation Evidence

Proveo — 11/11 PASS (v0.2.47, 2026-06-09T02:55Z)

Real Gmail sign-up (alembasic@gmail.com) end-to-end on `bilko-demo.alai.no`:

Step	Result	Details
1	PASS	Self-serve copy present; "Contact your administrator" absent
2	PASS	"Sign in with Microsoft" → ciamlogin.com (tenant 20bb17de) redirect
3	PASS	Email entered on CIAM; OTP sent immediately
4	PASS	Returning user — OTP sent directly (no create-account needed)
5	PASS	8-digit OTP (17717965) received via Gmail UID:75644 in 7 seconds
6	PASS	Redirect back to bilko-demo.alai.no/dashboard
7	PASS	POST /auth/entra/session → 200, Bilko HMAC JWT, org 4e96b6ff confirmed
8	PASS	/dashboard with trial UI ("Probno: 6 dana preostalo"), /auth/me → 200 + trialEndsAt 2026-06-15
9	PASS	/invoices via SPA nav — empty org (0 invoices), session alive
10	PASS	/invoices/new — invoice form visible, trial tenant usable
11	PASS	Regression clear — admin wall absent, self-serve copy confirmed

Zero /auth/refresh calls during SPA navigation after B1.3 fix (confirmed by network capture count=0). Cross-tenant RLS: org 4e96b6ff shows 0 invoices and 0 BAM balances (no data from other tenants).

Securion — LAUNCH WITH CONDITIONS

- CRITICAL: None found.
- HIGH: H1 (JIT provisioning rate gate) — must fix before scale launch (MC #103245).
- PASS areas: RS256 JWKS verification, issuer/audience pinning, OID as identity anchor, alg:none bypass blocked, org_id derived from DB (not Entra token), RLS fail-closed, FORCE RLS on all 9 tenant tables, role=viewer hardcoded (no self-escalation), trial re-signup

blocked, refresh token rotation (jti-based single-use), legacy auth endpoints retired (HTTP 410).

---

8. DEPLOY-MAP Reference

The CIAM substitutions live in `infrastructure/gcp/cloudbuild.yaml` under the `bilko-main-deploy` trigger. The Cloudflare Turnstile entries in DEPLOY-MAP.md cover the marketing landing forms and are unrelated to the CIAM auth flow. No DEPLOY-MAP.md changes are required for the CIAM self-serve trial feature — the trigger substitutions are already updated.

Do not add CIAM secrets to the DEPLOY-MAP secrets table — these are build-time substitutions injected directly from the trigger, not GCP Secret Manager secrets.

---

9. Environment Variables Reference

Variable	Service	Correct value note
NEXT_PUBLIC_ENTRA_CLIENT_ID	bilko-web-demo (build-time)	c2902239-ea63-41bd-8619-6cf096d7d45a (SPA app)
NEXT_PUBLIC_ENTRA_AUTHORITY	bilko-web-demo (build-time)	https://[tenant-id].ciamlogin.com/[tenant-id]/v2.0 — no user flow suffix needed
NEXT_PUBLIC_ENTRA_SCOPE	bilko-web-demo (build-time)	api://fe39e0f5.../access_as_user
ENTRA_EXTERNAL_ID_ISSUER	bilko-api-demo	https://[tenant-id].ciamlogin.com/[tenant-id]/v2.0
ENTRA_EXTERNAL_ID_AUDIENCE	bilko-api-demo	**c2902239-ea63-41bd-8619-6cf096d7d45a** (SPA client ID — NOT the API resource ID)
ENTRA_EXTERNAL_ID_JWKS_URL	bilko-api-demo	https://[tenant-id].ciamlogin.com/[tenant-id]/discovery/v2.0/keys

Critical: `ENTRA_EXTERNAL_ID_AUDIENCE` must be the SPA client ID (`c2902239`), not the API resource app ID. MSAL requests id_tokens with the SPA client as audience. If set to the API app ID, the backend rejects every CIAM idToken with audience mismatch.

Page created by Skillforge (MC #103232 WS-D, 2026-06-09). Source evidence: /tmp/evidence-103232/. Validation: Proveo 11/11 PASS + Securion LAUNCH WITH CONDITIONS.

Revision #1
Created 2026-06-09 00:54:24 UTC by John

Updated 2026-06-09 00:54:24 UTC by John