Backend Specification

Version: 1.1 Date: 2025-09-13

Contents

1) Purpose & Scope

This document specifies the backend for a multi-tenant executive dashboard that ingests **only Cisco Umbrella** telemetry (Reports v2 + Investigate), materializes weekly gold marts, and exposes **Bubble-ready** APIs. Phase 1 covers Threats/KPIs/Heatmap/Top-N/Shadow-IT/RC Outdated/AI narrative.

Appendix A.2 lists **Phase 2** routes (controls funnel, policy simulation, visual analytics, and IR).

2) Architecture Overview

Pipeline: Ingestion → Enrichment (Umbrella Investigate) → Core (dims/facts) → Weekly Gold Marts → AI layer → Public APIs.

Xano building blocks:

- Background Tasks (crons) for each stream and nightly jobs.
- PostgreSQL (Xano) for RAW/CORE/MART/AI.
- API Groups: `/v1/umbrella/*`, `/v1/shadow-it/*`, `/v1/ai/*`, plus `/v1/infra/*` primitives.
- Env Vars for secrets; middleware for **tenant guard** and **ETag**.

Time & locale: store **UTC**; present **Europe/Madrid**; weekly keys use ISO Week (Mon–Sun).

3) Non-Functional Requirements (SLOs)

- **Latency:** P95 < **500 ms** for mart endpoints; P95 < **1.5 s** for heavy Top-N/detail joins.
- **Freshness:** hourly for 15-min/daily facts; weekly marts materialized nightly; current-week trend/heatmap/toplists refreshed hourly.
- **Availability:** ≥ 99.5% for public read APIs.
- **Retention:** facts **90 days**; marts **24 months** (per-tenant configurable).

4) Data Model (Summary)

Layers:

- **RAW (Bronze):** 1:1 Umbrella payloads, `_hash`, `schema_version`, `ingested_at`, `src_batch_id`, `natural_id` for idempotent upsert.

- **CORE (Silver):** `dim_time_utc`, `dim_identity` (SCD2), `dim_domain`, `dim_category`, optional `dim_app_saas`, `dim_cdfw_rule`; facts (DNS 15m/daily, CASB daily, optional CDFW 15m/daily, RC health daily, Investigate enrich daily, IOC events, domain relations).

- **MART (Gold, weekly):** `weekly_kpis_umbrella`, `risk_semaphore_weekly`, trend/evolution/heatmap/toplists/non-security categories, Shadow-IT flags & top apps, RC outdated, **Advanced Detections**.

- **AI (Governed):** `ai.baselines`, `ai.insights`, `ai.recommendations`, `ai.weekly_exec`, `ai.playbooks`, `ai.feedback`.

Keys & indexing: BRIN on time columns in high-volume facts; composite indexes `(tenant_id, iso_year, iso_week)` on marts; SCD2 dims with unique natural keys.

5) ETL/ELT & Scheduling

Crons (hourly):

- `ingest_dns_activity` (sliding window 24–48 h, paginate until empty) → `raw_dns_activity`.
- `ingest_identities` / `ingest_roaming_clients` → RAW.
- `ingest_casb_daily` → `raw_casb_app_usage`.
- `ingest_cdfw_events` (optional) → RAW → rollups 15m/daily.
- `rollup_transforms` → aggregate to CORE 15m/daily; current-week toplists/heatmap.
- `ai_baselines_anomalies` → recompute rolling baselines, flag outliers.

Nightly:

- `investigate_enrich_domains` (POST batch ≤1000 domains, cache 24 h) → RAW Investigate tables; rate-limit with exponential backoff.
- `materialize_weekly_marts` → build all gold marts.
- `T-24h remediation` for late arrivals (daily) and `T-2h` for 15m.

Idempotency & DLQ: UPSERT by `(tenant_id, natural_id)` with `_hash`; dead-letter queue for malformed/failed records.

6) API Design (Public)

Base path: `/v1`

Auth: Bearer JWT or API Key (header). All requests must resolve to a **tenant_id**.

Multitenancy guard: middleware injects `tenant_id` filter into every query.

Response envelope (lists):

```
{ "items": [ ... ], "meta": { "count": 123, "page": 1, "page_size": 20,
"next": 2 } }
```

Caching: `ETag/If-None-Match` where `ETag = hash(tenant_id + query_params + latest_updated_at)`. TTL: 60–300 s. Return **304** if unchanged.

Pagination: `page`, `page_size` (default 20, max 100).

Errors: `{ "code": "string", "message": "string" }` with appropriate HTTP status.

Filtering & params: use `tenant_id` + `iso_year` + `iso_week` for weekly reads; optional filters (`family`, `identity_id`, `limit`, etc.) per route.

7) Public API (Phase 1)

7.1 KPIs & Risk

- **GET** `/v1/umbrella/kpis-weekly?tenant_id&iso_year&iso_week`
  - Returns KPI cards (e.g., block rate %, TLS inspection %, agent coverage %, GRI, high-risk destinations, Shadow-IT KPIs). When licensing data is absent, any KPI that requires seats/licensing is omitted or returns `null`.
- **GET** `/v1/umbrella/risk-semaphore?tenant_id&iso_year&iso_week`
  - Malware/Phishing/C2/Cryptomining levels and RC outdated level (LOW/MED/HIGH/CRIT).

7.2 Trends & Heatmap

- **GET** `/v1/umbrella/trend-critical-4w?tenant_id`
- **GET** `/v1/umbrella/weekly-evolution?tenant_id&iso_year&iso_week`
- **GET** `/v1/umbrella/heatmap?tenant_id&iso_year&iso_week`

7.3 Top Lists & Categories

- **GET** `/v1/umbrella/top-identities?tenant_id&iso_year&iso_week&family=*&limit=10`
- **GET** `/v1/umbrella/top-domains?tenant_id&iso_year&iso_week&family=*&limit=10`
- **GET** `/v1/umbrella/nonsec-categories/top?tenant_id&iso_year&iso_week&limit=10`

7.4 Shadow-IT & Infra/RC Health

- **GET** `/v1/shadow-it/flags?tenant_id&iso_year&iso_week`
- **GET** `/v1/shadow-it/top-apps?tenant_id&iso_year&iso_week&limit=20`
- **GET** `/v1/umbrella/rc/outdated?tenant_id&iso_year&iso_week`
- **GET** `/v1/umbrella/infra/status?tenant_id`
  - **Unified** to only `tenant_id` (current heartbeat and connector/tunnel health). No week params.

- **GET** `/v1/umbrella/licensing?tenant_id`
  - Administrative snapshot to support KPIs that need a denominator.
  - **Response (example):**

```
{ "items": [{
    "seats_total": 1200,
    "seats_assigned": 980,
    "seats_active_7d": 945,
    "license_tier": "SIG Essentials",
    "license_expiry": "2026-03-31"
}], "meta": {"tenant_id": "..."} }
```

  - **Derived KPIs:**
    - `agent_coverage_pct = seats_active_7d / seats_assigned` (if assigned>0)
    - `license_utilization_pct = seats_assigned / seats_total` (if total>0)

7.5 AI (Narrative & Recommendations)

- **GET** `/v1/ai/insights?tenant_id&from&to&severity=HIGH,CRITICAL&page=1&page_size=50`
- **GET** `/v1/ai/recommendations?tenant_id&iso_year&iso_week`
- **GET** `/v1/ai/weekly-exec?tenant_id&iso_year&iso_week`

Contracts: All return the standard list envelope. For enrich-dependent bullets, include `enrichment_last_updated` so the UI can display a **stale badge** when age > 24 h.

8) Security & Multitenancy

- Every table is keyed by `tenant_id`; every endpoint enforces a **tenant guard** in middleware.
- Secrets (Umbrella, Investigate) in environment variables; rotate and scope least privilege.
- **PII minimization:** prefer Umbrella identity labels; avoid storing user emails beyond what Umbrella provides; hash WHOIS emails from Investigate.
- **Auditing & rate limits:** log `audit_api_calls` with params hash; throttle per tenant; circuit breaker on repeated Cisco 429/5xx.
- **Defense-in-depth (recommended):** add **Row-Level Security (RLS)** policies on PostgreSQL if available in Xano's plan.

9) Observability & Data Quality

- **Metrics:** rows/sec, lag, duplicates %, error rate per stream; store per tenant and stream.
- **DQ rules:** totals reconciliation (allowed+blocked=total), identity/domain cardinalities, RC active coverage %, outlier detection (same-DOW median), and freshness checks for Investigate enrichment.
- **Checkpoints:** per-stream watermarks; daily `T-24 h` recompute for late data.

10) Performance, Indexing & Retention

- **Facts:** BRIN on `ts_15m`; composite `(tenant_id, ts_15m)`; retain 90 d.
- **Dailies & Marts:** `(tenant_id, date_id)` or `(tenant_id, iso_year, iso_week)`; retain 24 m.
- **Dims:** unique indexes on natural keys; `dim_identity` as SCD2.
- **Compatibility note:** if native partitioning/UNLOGGED is unavailable, keep BRIN + scheduled pruning.

11) Error Handling & Runbooks

- **401/403:** refresh tokens; verify tenant mapping; rotate secrets if needed.
- **429 (Cisco):** exponential backoff with jitter; shrink window; pause per stream.
- **5xx (Cisco):** retry with backoff; open circuit if sustained.
- **Backfills:** parameterized date-chunk jobs; ensure DLQ is drained; replay with idempotent UPSERT.
- **Investigate:** POST batch ≤1000; dedupe by domain/day; cache 24 h; tag enrich timestamps.

12) Environments & Deployment

- Environments: **prod**
- CI/CD: versioned DB migrations; feature flags per tenant; private webhook (HMAC) to invalidate Bubble caches after marts complete.

13) Field-to-Widget Mapping (Gold Marts → UI)

- Executive KPIs ↔ `mart.weekly_kpis_umbrella` (plus licensing snapshot when present).
- Risk Semaphore ↔ `mart.risk_semaphore_weekly`.
- 4-week Trend ↔ `mart.trend_critical_blocks_4w`.
- Weekly Evolution (DOW) ↔ `mart.weekly_evolution_blocks`.
- Heatmap 7×24 ↔ `mart.heatmap_hourly_week`.

- Top Identities / Domains ↔ `mart.top_identities_weekly` / `mart.top_domains_weekly`.
- Non-security Categories ↔ `mart.nonsec_block_categories_weekly`.
- Shadow-IT ↔ `mart.shadowit_flags_weekly`, `mart.shadowit_top_apps_weekly`.
- RC Outdated ↔ `mart.rc_outdated_weekly`.
- Advanced Detections ↔ `mart.advanced_detections_weekly`.
- Weekly Narrative ↔ `ai.weekly_exec`; Insights/Recommendations ↔ `ai.insights`, `ai.recommendations`.

14) Acceptance Criteria

1) KPI cards show value + ΔWoW + sparkline; tooltips include formula & source; GRI matches materialized value.
2) Risk semaphore & RC outdated render levels per server payload.
3) Top-N, trend, heatmap, categories, Shadow-IT, and RC Outdated reconcile ±1% with marts for the selected week.
4) AI narrative and recommendations populate Executive sections; evidence links open the correct filtered views.
5) ETag honored (304 on unchanged) with TTL 60–300 s; list envelope keys present.
6) Investigate-dependent widgets display **stale badge** when `enrichment_last_updated` > 24 h.

Appendix A.1 — Full List of Public Endpoints (Phase 1)

- `/v1/umbrella/kpis-weekly`
- `/v1/umbrella/risk-semaphore`
- `/v1/umbrella/trend-critical-4w`
- `/v1/umbrella/weekly-evolution`
- `/v1/umbrella/heatmap`
- `/v1/umbrella/top-identities`
- `/v1/umbrella/top-domains`
- `/v1/umbrella/nonsec-categories/top`
- `/v1/shadow-it/flags`
- `/v1/shadow-it/top-apps`
- `/v1/umbrella/rc/outdated`
- `/v1/umbrella/infra/status` (only `tenant_id`)
- `/v1/umbrella/licensing`
- `/v1/ai/insights`
- `/v1/ai/recommendations`
- `/v1/ai/weekly-exec`

Standard contracts: list envelope + ETag + tenant guard.

Appendix A.2 — Phase 2

Controls & Policy: - `/v1/umbrella/controls-funnel?tenant_id&iso_year&iso_week` — end-to-end funnel (total → policy hits → overrides → noise filtered). - `/v1/umbrella/policy-simulation?tenant_id&iso_year&iso_week&sim=RISK_GE_80|NEWLY_SEEN|...` — reads `mart.policy_simulation_weekly`.

Visual Analytics: - `/v1/vis/sankey?tenant_id&iso_year&iso_week&min_flow=50&max_nodes=80` — from `mart.flow_identity_category_verdict_weekly`. - `/v1/vis/sunburst?tenant_id&iso_year&iso_week&limit_per_category=25` — from `mart.sunburst_category_domain_weekly`.

SWG/CDFW Supporting Marts & Routes: - `/v1/umbrella/swg/traffic` (DOW allowed/blocked), `/v1/umbrella/swg/tls-coverage`, `/v1/umbrella/swg/latency-outliers`. - `/v1/umbrella/cdfw/blocked-sessions` (group_by facet), `/v1/umbrella/cdfw/geo` (map), `/v1/umbrella/tunnels/health`.

Incident Response (IR) — Optional local incidents: - Weekly marts: `mart.ir_incidents_weekly`, `mart.ir_sla_weekly`. - **Severity enum (normalized):** `CRITICAL|HIGH|MEDIUM|LOW` (replace any legacy Q2). - Potential endpoints: `/v1/ir/incidents?tenant_id&iso_year&iso_week&severity=*`, `/v1/ir/sla?tenant_id&iso_year&iso_week`.

Benchmarks & Policy Diff (optional): - `/v1/ai/policy-diff?tenant_id&policy_key&from&to`. - `/v1/ai/benchmarks?tenant_id&industry=*`.

Appendix B — Example Payloads

/v1/umbrella/top-domains (simplified)


```
{
  "items": [
    { "domain": "evil.example", "family": "commandandcontrol", "blocks":
1240,
      "delta_wow_pct": 85.3, "risk_score": 92.0,
      "whois_created": "2025-07-02", "asn_name": "AS13335", "asn_number":
13335,
      "impacted_identities": 37 }
  ],
  "meta": {"count": 20}
}
```


/v1/ai/recommendations (simplified)


```
{
  "items": [
```

```
    { "reco_id": "reco-8f7a", "title": "Block domain auth-m365-portal.net
across all policies",
      "priority": "High", "impact": "High", "effort": "Low", "owner": "IT
Security",
      "eta": "2025-08-23", "status": "Not Started",
      "evidence_links": ["/threats?family=phishing&week=2025-W33&domain=auth-
m365-portal.net"],
      "tags": ["phishing", "policy"] }
  ],
  "meta": {"tenant_id": "...", "iso_year": 2025, "iso_week": 33}
}
```

Appendix C — KPI/Metric Notes

- **Block Rate %** = `security_blocks / total_requests`.
- **TLS Inspection %** = `inspected_swg_requests / total_swg_requests`.
- **Agent Coverage %** = `seats_active_7d / seats_assigned` (requires `/licensing`).
- **License Utilization %** = `seats_assigned / seats_total` (requires `/licensing`).
- **Global Risk Index (GRI):** weighted composite of threat severity, identity risk, and Shadow-IT exposure (tenant-overridable thresholds/weights in settings).

Appendix D — Implementation Hints (Xano)

- Precompute weekly metrics in `mart.weekly_kpis_umbrella`; deltas in `mart.exec_delta_weekly`; narrative in `ai.weekly_exec`.
- Compute `ETag` server-side; honor `If-None-Match`.
- If RLS is available, enforce row policies on marts/facts keyed by `tenant_id`.
- If partitioning/UNLOGGED is not available, retain BRIN + scheduled pruning.
- Provide a private cache-bust webhook (HMAC) after `materialize_weekly_marts` completes.