

## GOVERNED TEMPLATE RENDERING REFERENCE: INFRASTRUCTURE & DEPLOYMENT DEFAULT V2 (DRAFT)

definition 23d6a3f8...ed06

INFRASTRUCTURE & DEPLOYMENT · PROJECT ARTIFACT

SPECIRA

# Infrastructure & Deployment: Dispatch Modernization

Meridian Field Services: instructional example, not project evidence

DRAFT · WATERMARK POLICY: DRAFT\_ONLY

template infrastructure\_deployment v2 · pack: specira\_default\_delivery

the architecture owns topology; this artifact owns the operational realization

## §1 Hosting Strategy

mandatory

1 decision

1 evidence rule

validators: hosting\_decision\_logged\_with\_alternatives · re\_decision\_triggers\_are\_conditions · every\_secret\_sourced\_from\_config\_not\_code

Decision: managed container platform

**What:** a managed PaaS runs the four services; cloud primitives hand the team undifferentiated plumbing; Kubernetes adds an orchestration team the project does not have. **Why considered:** four services, one team, pilot scale; nothing scales heterogeneously.

**Who & how:** E. Sandoval with the delivery lead, walked against the container view July 17 ([hosting decision<sup>\[1\]</sup>](#)).

**Re-decision triggers, stated:** ten or more services, per-service scaling profiles diverging, or a second team needing isolated deploys; none fire at pilot. **Secrets:** the platform secret store per the config law; machine credentials per the [machine-identity rows<sup>\[2\]</sup>](#); nothing in code.

## §2 Network Topology

mandatory

1 decision

1 evidence rule

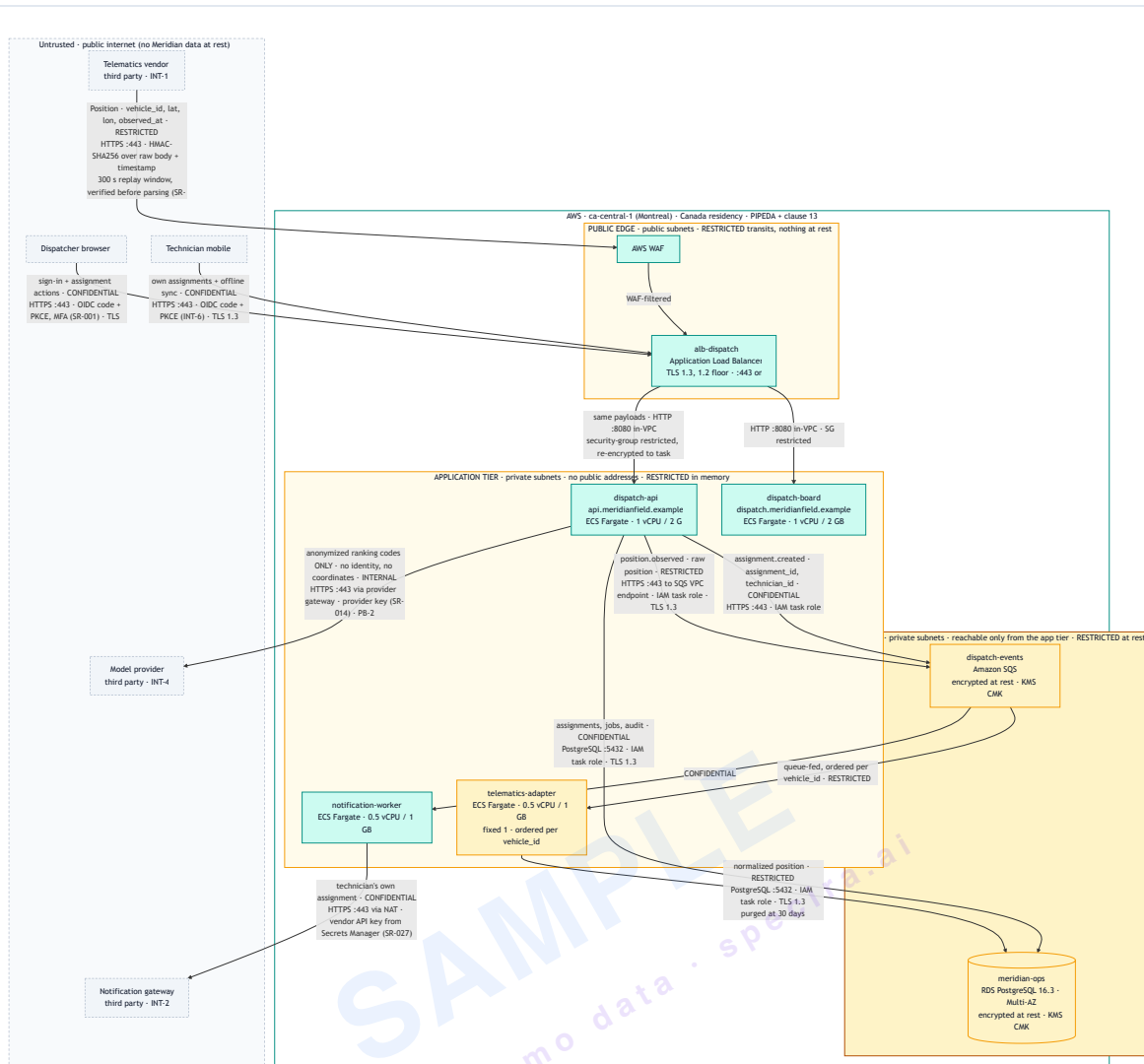
validators: provider\_and\_region\_named\_not\_only\_platform\_class ·  
 region\_carries\_residency\_justification ·  
 every\_node\_names\_auth\_mechanism\_citing\_security\_requirement ·  
 every\_edge\_names\_actual\_payload\_fields\_not\_generic\_data ·  
 assessor\_can\_evaluate\_topology\_without\_opening\_another\_artifact

§1 chose a class of platform. A class cannot be assessed, so this section names the actual one, and is the project's **infrastructure register**: the single place the concrete facts live, which every other artifact cites rather than keeping a second copy that drifts. The architecture owns the LOGICAL boundaries ([boundaries<sup>\[6\]</sup>](#)); this realizes them.

#### DECISION: AMAZON WEB SERVICES, REGION CA-CENTRAL-1 (MONTREAL)

one account per environment; the managed-container-platform class from §1 is realized as Elastic Container Service on Fargate (no cluster hosts to patch, the whole point of the class). *Residency justification*: technician location is personal data that must stay in Canada (PIPEDA and enterprise contract clause 13, [SR-033<sup>\[11\]</sup>](#); [handling matrix<sup>\[10\]</sup>](#)). ca-central-1 is a Canadian region and processing is single-region, so the rule is satisfied **by placement rather than by policy**: the strongest form, because it cannot be forgotten. *Who*: E. Sandoval with N. Duval (compliance) ([provider and region<sup>\[1\]</sup>](#)).

SAMPLE  
 seeded demo data · specira.ai



Three trust zones inside ca-central-1, shaded by the highest data class each holds. Fallback text: the public internet is untrusted and holds no Meridian data at rest. The PUBLIC EDGE (public subnets) contains AWS WAF and the application load balancer, the only internet-reachable construct, listening on 443 only and terminating TLS 1.3 with a 1.2 floor; RESTRICTED data transits it but nothing rests there. The APPLICATION TIER (private subnets, no public addresses) runs the four Fargate services: dispatch-api, dispatch-board, telematics-adaptor, and notification-worker. The DATA TIER (private subnets, reachable only from the application tier) holds the meridian-ops RDS PostgreSQL 16.3 Multi-AZ instance and the dispatch-events SQS queue, both encrypted at rest with a customer-managed KMS key. PB-1 is realized as the load balancer's webhook listener; PB-2 as the single provider-gateway egress path. The node and edge tables below carry the full detail.

## Node table: what each box actually is

| Node · hostname                                  | Runs on                                                       | Callers authenticate with                                                                                                          | Encryption (transit · rest)           | Highest class held                  |
|--------------------------------------------------|---------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------|-------------------------------------|
| alb-dispatch                                     | Application Load Balancer · public subnets · AWS WAF in front | n/a (terminates TLS; 443 only)                                                                                                     | TLS 1.3 (1.2 floor) · nothing at rest | <b>RESTRICTED</b> transits          |
| dispatch-api<br>api.meridianfield.example        | ECS Fargate · application tier · 1 vCPU / 2 GB                | Meridian SSO, OIDC code + PKCE (SR-001 <sup>[11]</sup> ); the webhook route instead verifies HMAC-SHA256 (SR-021 <sup>[11]</sup> ) | TLS 1.3 · no local persistence        | <b>CONFIDENTIAL</b> in flight       |
| dispatch-board<br>dispatch.meridianfield.example | ECS Fargate · application tier · 1 vCPU / 2 GB                | Meridian SSO with MFA (passkey), required for operations-manager (identity model <sup>[2]</sup> )                                  | TLS 1.3 · holds nothing               | n/a (serves to browser application) |
| telematics-adapter                               | ECS Fargate · application tier · 0.5 vCPU / 1 GB · fixed 1    | <b>No ingress listener:</b> queue-fed, reachable from nothing                                                                      | TLS 1.3 to queue and RDS · none local | <b>RESTRICTED</b> raw position      |
| notification-worker                              | ECS Fargate · application tier · 0.5                          | <b>No ingress listener:</b> queue-fed; egresses                                                                                    | TLS 1.3 · none local                  | <b>CONFIDENTIAL</b>                 |

| Node · hostname | Runs on                                    | Callers authenticate with                                              | Encryption (transit · rest)        | Highest class held                       |
|-----------------|--------------------------------------------|------------------------------------------------------------------------|------------------------------------|------------------------------------------|
|                 | vCPU / 1 GB                                | with the vendor API key from Secrets Manager (SR-027 <sup>[11]</sup> ) |                                    |                                          |
| meridian-ops    | RDS PostgreSQL 16.3 · Multi-AZ · data tier | IAM task roles only; no human credential path                          | TLS 1.3 · KMS customer-managed key | <b>RESTRICTED</b><br>the positions table |
| dispatch-events | Amazon SQS · data tier · VPC endpoint      | IAM task roles only                                                    | TLS 1.3 · KMS customer-managed key | <b>RESTRICTED</b>                        |

Edge table: what actually moves on each arrow

| Edge                         | Payload (the actual fields)                |
|------------------------------|--------------------------------------------|
| vendor → alb-dispatch (PB-1) | Position.vehicle_id, lat, lon, observed_at |
| alb → dispatch-api           | the same payload, in-VPC                   |

| Edge                                           | Payload (the actual fields)                                                              |
|------------------------------------------------|------------------------------------------------------------------------------------------|
|                                                |                                                                                          |
| dispatch-api → dispatch-events                 | position.observed: the raw position                                                      |
| telematics-adapter → meridian-ops              | the normalized position; <b>purged at 30 days</b><br>( <u>retention<sup>[10]</sup></u> ) |
| dispatch-api → dispatch-events                 | assignment.created: assignment_id, technician_id                                         |
| notification-worker → gateway<br>(INT-2)       | the technician's own assignment                                                          |
| dispatch-api → model provider<br>(PB-2, INT-4) | <b>anonymized ranking codes ONLY:</b> no identity, no coordinates                        |

|      |                             |
|------|-----------------------------|
| Edge | Payload (the actual fields) |
|------|-----------------------------|

**Why PB-2 carries INTERNAL and not RESTRICTED:** the anonymization happens *before* the crossing, so no personal data is ever presented to the provider (sensitive data paths<sup>[12]</sup>). This is a placement property, not a promise.

INGRESS AND EGRESS: DENY BY DEFAULT, BOTH DIRECTIONS

*Ingress:* only 443 to the load balancer, from the internet, WAF-filtered; nothing else is reachable, and the application and data tiers have no public addresses. *Egress:* exactly three destinations are allowed, each justified, namely the notification gateway (INT-2, assignment delivery), the provider gateway (INT-4, ranking), and the AWS service endpoints. Everything else is refused at the security group. Each exception is owned by E. Sandoval.

**The cold-read test:** a reviewer should be able to answer the only question that matters, *where does RESTRICTED data rest and move?*, from this section alone: it enters at the load balancer's webhook listener under HMAC verification, transits dispatch-api and the queue, is normalized by the adapter, rests in the positions table under a customer-managed KMS key, is hard-purged at 30 days, and **never crosses PB-2 and never leaves ca-central-1**. Backups go to ca-west-1 (Calgary), a second Canadian region, so isolation and residency hold together (§6).

### §3 Environment Strategy

mandatory

1 decision

1 evidence rule

validators: every\_environment\_has\_purpose\_and\_parity\_note · every\_parity\_gap\_has\_named\_owner · drift\_detection\_scheduled\_and\_alerting

*Each environment states what it PROVES; every unclosed parity gap has a named owner. Console changes are drift by definition.*

| Environment | Proves                 | Data source                                 | Parity note                                                      | Owner · cost tag    |
|-------------|------------------------|---------------------------------------------|------------------------------------------------------------------|---------------------|
| dev         | Isolated iteration     | Synthetic ( <u>seeding</u> <sup>[3]</sup> ) | Shared platform tier: gap accepted; no load claims made from dev | Delivery lead · dev |
| staging     | Production-shaped end- | Masked ( <u>seeding</u> <sup>[3]</sup> )    | Database <b>one tier below</b> production sizing, the named      | E. Sandoval ·       |

| Environment | Proves                 | Data source | Parity note                                                                             | Owner · cost tag   |
|-------------|------------------------|-------------|-----------------------------------------------------------------------------------------|--------------------|
|             | to-end behavior        |             | gap, owner E. Sandoval; limits the absolute-ceiling claim (environment <sup>[4]</sup> ) | staging            |
| production  | Real scale, real users | Real        | n/a                                                                                     | E. Sandoval · prod |

**laC law:** all three reproducible from the infrastructure repository; drift check nightly with alerting to the platform channel.

§4 Container & Orchestration

mandatory

1 decision

1 evidence rule

validator: orchestration\_only\_if\_earned\_statement\_present

**Realization:** four service images (API, board, adapter, worker) built once and promoted by reference (pipeline stages<sup>[5]</sup>). Runtime limits per class: API and board 1 vCPU / 2 GB; adapter and worker 0.5 vCPU / 1 GB, from the pilot frame (scale frame<sup>[6]</sup>). **Only-if-earned:** the platform owns scheduling, restarts, and networking; the project sits well under every orchestration trigger from §1, stated plainly.

§5 Scaling Strategy

mandatory

1 decision

1 evidence rule

validators: scaling\_numbers\_cited\_never\_rederived · stateful\_constraints\_cite\_architecture\_decision

Every number is a citation: the frame lives in the NFR catalog and architecture; this section states only the HOW.

| Service class | Policy                                                                                | Cites                                                        |
|---------------|---------------------------------------------------------------------------------------|--------------------------------------------------------------|
| API · board   | Autoscale on request concurrency, 1 to 3 instances (pilot headroom)                   | scalability rows <sup>[7]</sup> · scale frame <sup>[6]</sup> |
| Adapter       | <b>Single-instance by design:</b> ordered feed reconciliation; scales vertically only | INT-1 design <sup>[6]</sup>                                  |
| Worker        | Autoscale on queue depth, 1 to 2 instances                                            | scale frame <sup>[6]</sup>                                   |

**Load shedding:** the API contract's rate limits ([limits<sup>\[8\]</sup>](#)). The fivefold breakpoint margin is the perf artifact's finding, cited ([PS-3 results<sup>\[4\]</sup>](#)), re-derived nowhere.

§6 Disaster Recovery

mandatory

1 decision

1 evidence rule

validators: dr\_numbers\_cite\_nfr\_catalog\_not\_invented · every\_dr\_row\_has\_drill\_history\_or\_first\_drill\_date

Objectives are cited; actuals are drilled. A number without a drill is paper.

| Data class           | RPO                                                                                                                                            | RTO | Mechanism                                      | Last drill · actual                                           |
|----------------------|------------------------------------------------------------------------------------------------------------------------------------------------|-----|------------------------------------------------|---------------------------------------------------------------|
| Assignment records   | zero                                                                                                                                           | 4 h | Continuous replication + point-in-time restore | Sep 9, 2026 · RTA 2 h 40 m vs 4 h ( <a href="#">restore</a> ) |
| Jobs & configuration | 24 h                                                                                                                                           | 4 h | Same restore path                              | Sep 9, 2026 · same drill                                      |
| Positions            | Irrelevant by design: the 30-day purge ( <a href="#">retention<sup>[10]</sup></a> ) outruns any restore need; stated so the absence is a claim |     | n/a                                            |                                                               |

Objectives cite [NFR-A3<sup>\[7\]</sup>](#). **Backups:** encrypted, platform-isolated from the primary region; retention follows the [classification rules<sup>\[10\]</sup>](#). **Drill cadence:** quarterly, timed, including application verification; the next objective-versus-actual gap escalates to the sponsor.

§7 Open Questions

mandatory

1 decision

| Question                                                                                                                                                                        | Owner       | Answer by    | Blocks                          |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|--------------|---------------------------------|
| Does the staging database tier close to production sizing before the hub-six breakpoint re-run (the named parity gap's revisit, <a href="#">open questions<sup>[4]</sup></a> )? | E. Sandoval | Sep 30, 2026 | The absolute-ceiling claim only |

Refs

References & Package Contents

In the Specira workspace

|         |                                                                                                                                                                                                                   |
|---------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| specira | [1] Decision log: hosting decision app.specira.ai/projects/dispatch-modernization/artifacts/decision-log                                                                                                          |
| specira | [2] Auth & authz policy: machine identity app.specira.ai/projects/dispatch-modernization/artifacts/auth-authz-policy#machine                                                                                      |
| specira | [3] Data migration: seeding rules per environment app.specira.ai/projects/dispatch-modernization/artifacts/data-migration#seeding                                                                                 |
| specira | [4] Performance test scenarios: environment parity, PS-3 results, open questions app.specira.ai/projects/dispatch-modernization/artifacts/performance-test-scenarios                                              |
| specira | [5] CI/CD pipeline: build-once-promote-many stages app.specira.ai/projects/dispatch-modernization/artifacts/cicd-pipeline#stages                                                                                  |
| specira | [6] Architecture: container view, scale frame, INT-1 single-writer design app.specira.ai/projects/dispatch-modernization/artifacts/architecture                                                                   |
| specira | [7] NFR catalog: scalability rows, NFR-A3 recovery objectives app.specira.ai/projects/dispatch-modernization/artifacts/nfr-catalog                                                                                |
| specira | [8] API contracts: rate limits app.specira.ai/projects/dispatch-modernization/artifacts/api-contracts#limits                                                                                                      |
| specira | [9] Deployment verification: restore check app.specira.ai/projects/dispatch-modernization/artifacts/deployment-verification                                                                                       |
| specira | [10] Data classification: the handling matrix; retention rules app.specira.ai/projects/dispatch-modernization/artifacts/data-classification#retention                                                             |
| specira | [11] Security requirements: SR-001 identity, SR-014 provider key, SR-021 feed authentication, SR-027 gateway key, SR-033 residency app.specira.ai/projects/dispatch-modernization/artifacts/security-requirements |
| specira | [12] Data flow diagrams: sensitive data paths; the PB-2 anonymization boundary app.specira.ai/projects/dispatch-modernization/artifacts/data-flow-diagrams#sensitive-paths                                        |