



Secure OT Log Collection with DataBahn Smart Edge

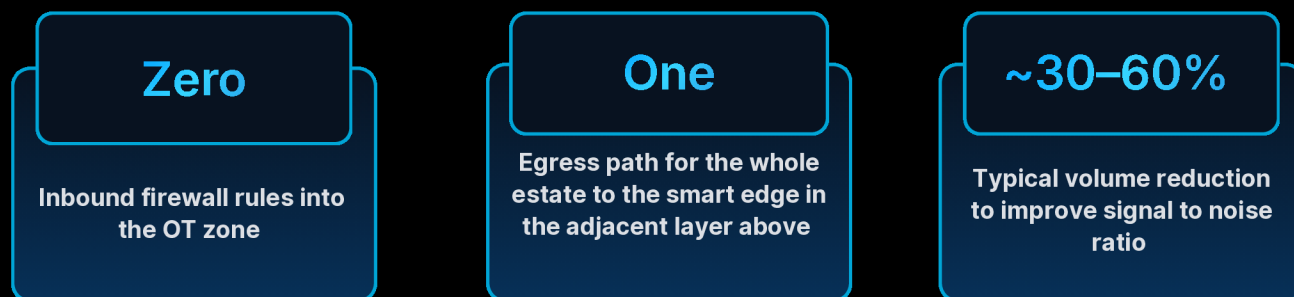
Collecting telemetry across Purdue Levels 0–5 without inbound firewall rules or internet access in the OT zone



AUTHORED BY

**Tejeswara Reddy
Ponnathota**

Senior Solutions Architect, DataBahn



Executive Summary

The Purdue Model is the standard blueprint industrial sites use to segment their networks. Physical equipment sits at the bottom, corporate IT at the top, with strict rules on what can talk to what.

Purdue-segmented OT environments are intentionally isolated from external networks and restrict inbound communications from higher-level systems. While essential for operational security, this isolation often prevents critical operational telemetry from reaching enterprise monitoring, SOC, and observability platforms.

DataBahn deploys Smart Edge nodes inside each zone and chains them: a node with no internet connectivity forwards to a DataBahn node in an adjacent zone that does. Each relay accepts connections only from the DataBahn node directly below it and initiates connections only to the DataBahn node directly in the above layer, and no other traffic is permitted through a proxy hop.

The result is a single outbound path from the deepest OT cell to outer most layer via relays before being delivered to the destination.

The Purdue Model

The Purdue Enterprise Reference Architecture is the reference segmentation model for industrial control systems, and the basis of the zone-and-conduit approach in IEC 62443.

It divides an industrial estate into hierarchical levels, from the physical process at the bottom to enterprise IT at the top, with an Industrial DMZ enforcing a hard break between OT and IT.

5	Enterprise Network	Corporate it, ERP, email, internet access, SOC and SIEM
4	Business Logistics	Site business planning, scheduling, plant IT services
IDMZ	Industrial DMZ	Enforces break between IT and IT - no direct traffic traverses this boundary
3	Site Operations	Historians, MES, batch and production management, OT domain services
2	Area Supervisory Control	HMI's, SCADA servers, engineering workstations, alarm systems
1	Basic Control	PLCs, RTUs, DCS controllers, safety instrumented systems
0	Physical Process	Sensors, actuators, drives, valves, field instrumentation

Figure 1: The Purdue model: hierarchical levels and the Industrial DMZ boundary.

What the model implies for log collection

- **Traffic is hierarchical, never end-to-end.** A device at Level 1 has no route to Level 5, and by design nothing at Level 0–3 has a route to the internet. Public DNS, NTP, and certificate revocation endpoints are typically unreachable too.
- **Inbound is the harder ask.** Even where an outbound rule can be justified, a rule allowing a higher level to open a session into a lower one is usually refused outright — it is the exact path an attacker would want.
- **Change control is slow and evidence driven.** Every firewall rule is reviewed against safety and availability impact, so a design needing many rules per source rarely gets deployed.

Any workable collection design therefore must satisfy three constraints at once: no inbound rules into a lower level, no internet dependency below the IDMZ, and as few firewall changes as possible. That is what the rest of this brief describes.

The Challenge

Industrial and critical-infrastructure environments are segmented according to the Purdue Enterprise Reference Architecture. Levels 0–2 (sensors, PLCs, HMIs, SCADA) and Level 3 (site operations, historians, MES) are isolated by design: no internet route, no public DNS, and no permitted inbound sessions from higher levels.

Security teams still need this telemetry for threat detection, insider-risk monitoring, and IEC 62443 / NERC CIP / NIS2 evidence.

Why the usual options fall short

- **Outbound holes from Level 2.** Punching internet egress from the process-control zone defeats the segmentation it is there to provide and is rarely approved by an OT change board.
- **Data diodes and one-way gateways.** Expensive per site, slow to change, and they typically move raw files rather than parsed, normalized events.
- **Per-zone SIEM collectors.** Multiply agent footprint, licensing, and the volume of low-value data shipped upstream.
- **Manual extraction to removable media.** Slow, error-prone, and useless for detection or response.

The DataBahn Solution

DataBahn deploys lightweight Smart Edge nodes, organized into a Fleet, inside each network zone. Each node collects locally using native connectors, processes data where it is generated, and forwards it upstream.

A node with no internet route doesn't need one: it hands its data to another DataBahn node in the zone above, using the DB Proxy Connector, the component that lets one node relay traffic on behalf of a node below it. Chaining these proxy hops produces a single, auditable egress path from the deepest OT cell to DataBahn Cloud, with no inbound rules and no new internet-facing listener at any level.

Three node roles

Role	Purdue placement	Function
Collection Node	Levels 0–2 — OT cell / process control	Terminates syslog, polls OPC-UA and Modbus gateways, tails files, reads local databases. Parses, filters, and buffers locally, then forwards to the relay node.
Relay Node	Level 3 — Site operations / private network	Collects Level 3 sources and simultaneously runs the DB Proxy Connector, accepting the TLS stream from the OT node and multiplexing it upstream.
Egress Node	Levels 4–5 — Corporate IT	Collects enterprise sources and acts as the single internet-facing hop: one outbound TLS 443 session to DataBahn Cloud.

The chain is arbitrarily deep

A four- or five-hop topology (cell → line → site DMZ → regional IT → corporate) is configured exactly like a two-hop one , with each node is told only its immediate upstream peer.

Reference Architecture

The diagram below shows a three-zone chain. Every arrow is an outbound-initiated session: no zone accepts a connection from a less-trusted zone, and only the Levels 4–5 node touches the internet.

Where a Smart Edge node isn't practical, such as directly on an engineering workstation or a legacy host, a lightweight agent can be deployed instead and forwards through the same relay chain, keeping the outbound-only, zone-to-zone pattern intact.

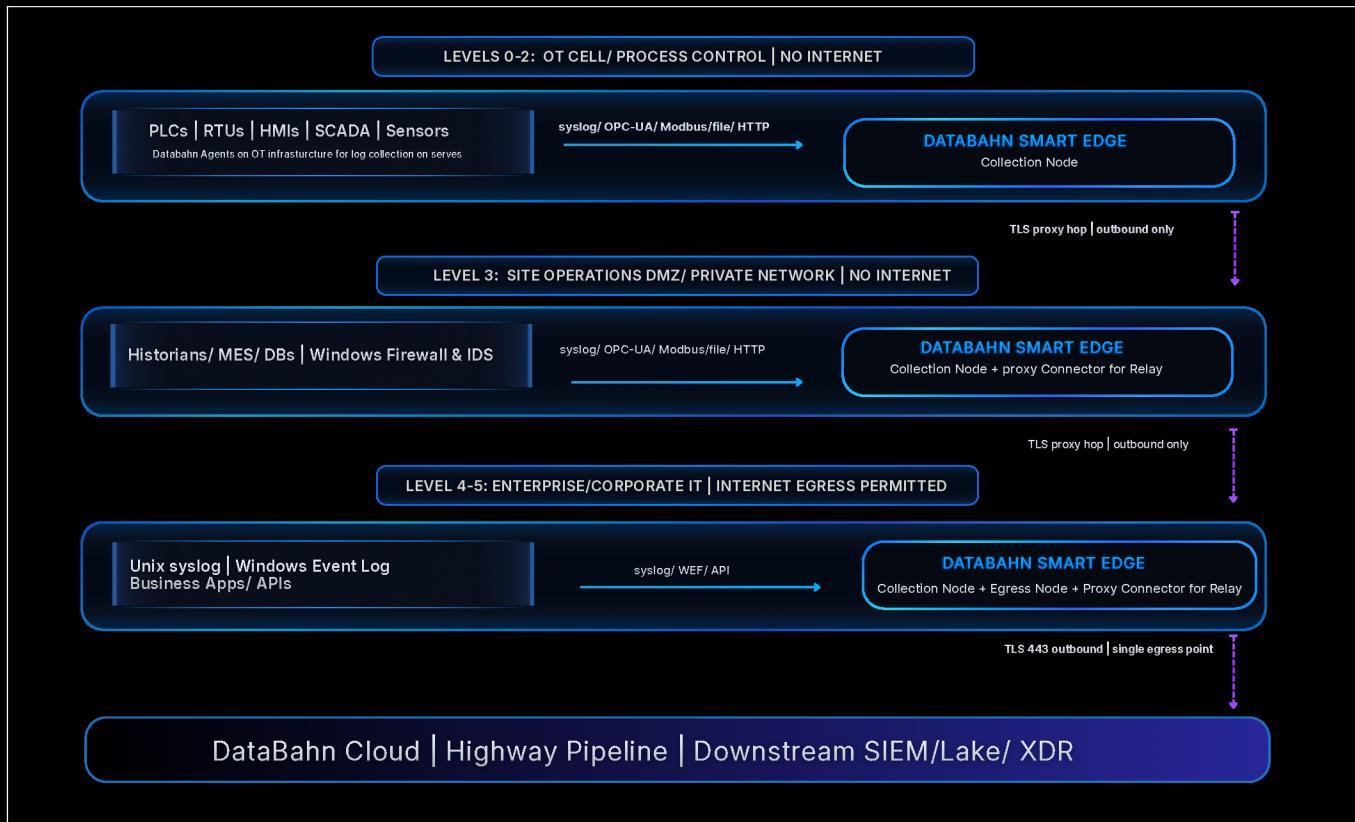


Figure 2: Chained DataBahn Smart Edge nodes across a Purdue-segmented OT environment

How Data is Collected

Every node runs the same connector library, so collection method is chosen per source rather than per zone.

Connector Class	Typical OT / IT sources	Mechanism
Syslog	Firewalls, switches, IDS/IPS (Clarity, Nozomi, Dragos), Linux/Unix hosts, network appliances	UDP/TCP/TLS listener on the in-zone node; source devices keep their existing destination config
Windows/agent	Engineering workstations, HMI servers, jump hosts, domain controllers	WEF/WEC subscription or lightweight agent reporting to the in-zone node
Database	Historians (PI, Ignition), MES, asset and CMMS databases	Scheduled JDBC/ODBC query that tracks the last record read, so it only pulls new rows
API/cloud	OT security platforms, vulnerability scanners, EAM, vendor SaaS	Polled REST/GraphQL with pagination and incremental cursors
File / OPC / MQTT / Net Flow/ OTEL / SNMP and more	PLC and RTU exports, batch reports, CSV/JSON drops, OPC-UA and Modbus gateways	Directory watch or protocol client on the collection node

Security and Network Controls

Connection model

- Outbound-only, per hop. Each node initiates the session to its upstream peer. Higher level nodes never initiate connections to lower level nodes. A DataBahn node at Level 1, for example, never opens a listener for Level 2 or Level 3 to connect into. It only reaches up. Within its own zone, the OT node exposes nothing beyond its local collection ports to the devices it's collecting from.
- Single egress point. Exactly one destination and one port leave the estate. TLS 443 routes to the DataBahn Cloud endpoint from the Levels 4–5 node. Deep zones need no internet, no proxy credentials, and no public DNS.
- TLS for secure transport. Every hop is TLS-encrypted and authenticated, so a rogue host inside a zone cannot register itself as a relay.
- Narrow, auditable rule set. A hop is one rule: source node IP, destination node IP, one TCP port. That is materially easier to get through OT change control than per-source internet egress.

Data integrity and resilience

- Built for low-bandwidth networks: DataBahn compresses every payload in transit, on each proxy hop and on egress, so constrained plant WAN, cellular, and satellite links carry a fraction of the raw volume, often the same 30-60% reduction seen upstream. Combined with edge filtering, only the events worth keeping reach the SOC.
- Persistent on-disk buffering at every node with configurable retention, so a WAN outage or upstream maintenance window causes queueing rather than loss.
- At-least-once delivery with checkpointed offsets and end-to-end acknowledgement across the full chain.
- Backpressure propagates downstream, a slow hop tells the hop below it to slow down too, so a saturated hop throttles collection instead of dropping events.
- Active/standby or clustered node pairs per zone for high availability; collection nodes fail over to an alternate relay peer.

Governance

- Central policy, connector configuration, and pipeline logic are authored in DataBahn Cloud and pulled by nodes over the same outbound channel , so nothing ever reaches back into the OT zone to push changes.
- Fleet health, throughput, buffer depth, and per-source volume are visible centrally, including for air-gapped nodes reporting through the chain.
- Signed configuration and node binaries; version pinning and staged rollout per zone to match OT maintenance windows.

Data Processing

Processing runs where the data is collected, which matters most on bandwidth-constrained plant networks. Sensitive fields never leave the zone that produced them.

- **Parse, normalize, and transform the data to OCSF/Sentinel/CIM/JSON**, so downstream tooling receives structured events regardless of source protocol.
- **Filter** low-value telemetry, typically ~30–60% volume reduction, directly lowering SIEM ingest cost.
- **Mask, tokenize, or drop** sensitive fields (operator identities, proprietary process values, PII) inside the zone, so regulated data never crosses the boundary.
- **Enrich and tag** with site, line, asset, and Purdue-level metadata for zone-aware detection, routing, and reporting.

Deployment Outline

Step	Activity
1	Zone and source inventory: map log sources to Purdue levels; identify which zones lack internet egress and where a relay is needed.
2	Sizing and Deployment: one node per OT cell or site, two for high availability, one relay per Level 3 DMZ, and an egress node pair in corporate IT.
3	Provision hosts: VM, container, or hardened appliance; no dependency on internet access for non-egress nodes.
4	Firewall change requests: one node-to-node rule per hop, plus a single TLS 443 egress rule to *.databahn.app from the egress node.
5	Configure and apply filtering, masking, and routing policy; measure volume reduction.
5	Route to destinations (SIEM, Data lake, SOAR, XDR, Observability platforms)

Outcomes

- Visibility without weakening segmentation. Level 0–3 telemetry reaches the SOC while the OT zone keeps zero internet exposure and zero inbound rules.
- One egress path to defend. A single, monitored, certificate-authenticated exit replaces a sprawl of per-collector internet permissions.
- Lower ingest and bandwidth cost. Filtering and normalization help reduce what crosses constrained plant links and what is billed downstream.
- Faster site rollout. A new site or cell is a node plus one firewall rule, reusing the existing policy set, not a new integration project.
- Audit-ready evidence. Zone-tagged, normalized, retained telemetry supports IEC 62443, NERC CIP, and NIS2 monitoring and logging requirements.
- Multi-destination support. The same collected stream can be routed to a SIEM, a data lake, and cold storage in parallel.

DataBahn is the agentic data control plane for the enterprise. Its platform ingests, normalizes, enriches, governs and routes telemetry from 700+ sources with complete data and AI neutrality: any source, any destination, any model, no vendor lock-in, at Fortune 100 scale. DataBahn is headquartered in Dallas and serves Fortune 500 enterprises.

Learn more at databahn.ai