DataBahn and Databricks

Next-Gen Security Data Infrastructure to Convert Telemetry into Intelligence

DATABAHN

DataBahn + Databricks

In cybersecurity, intelligence is the true differentiator – and it is only as strong as the data foundation beneath it. SOCs are expected to defend networks and business-critical data at AI speed, but they are overwhelmed by redundant alerts, siloed systems, and petabytes of raw telemetry that create bigger problems than the solution they are supposed to offer.

The solution is to place data at the foundation for agentic, AI-powered cybersecurity. This requires a shift in how the industry must think about security data: not as exhaust to be stored or queried, but as a living fabric that can be structured, enriched, and made ready for AI-native defense.

DataBahn and Databricks together make that shift real. By unifying fragmented telemetry and delivering it as trusted, actionable insight, we help enterprises move beyond log management into AI-powered security intelligence at scale.

Rethinking “more data = better defense”

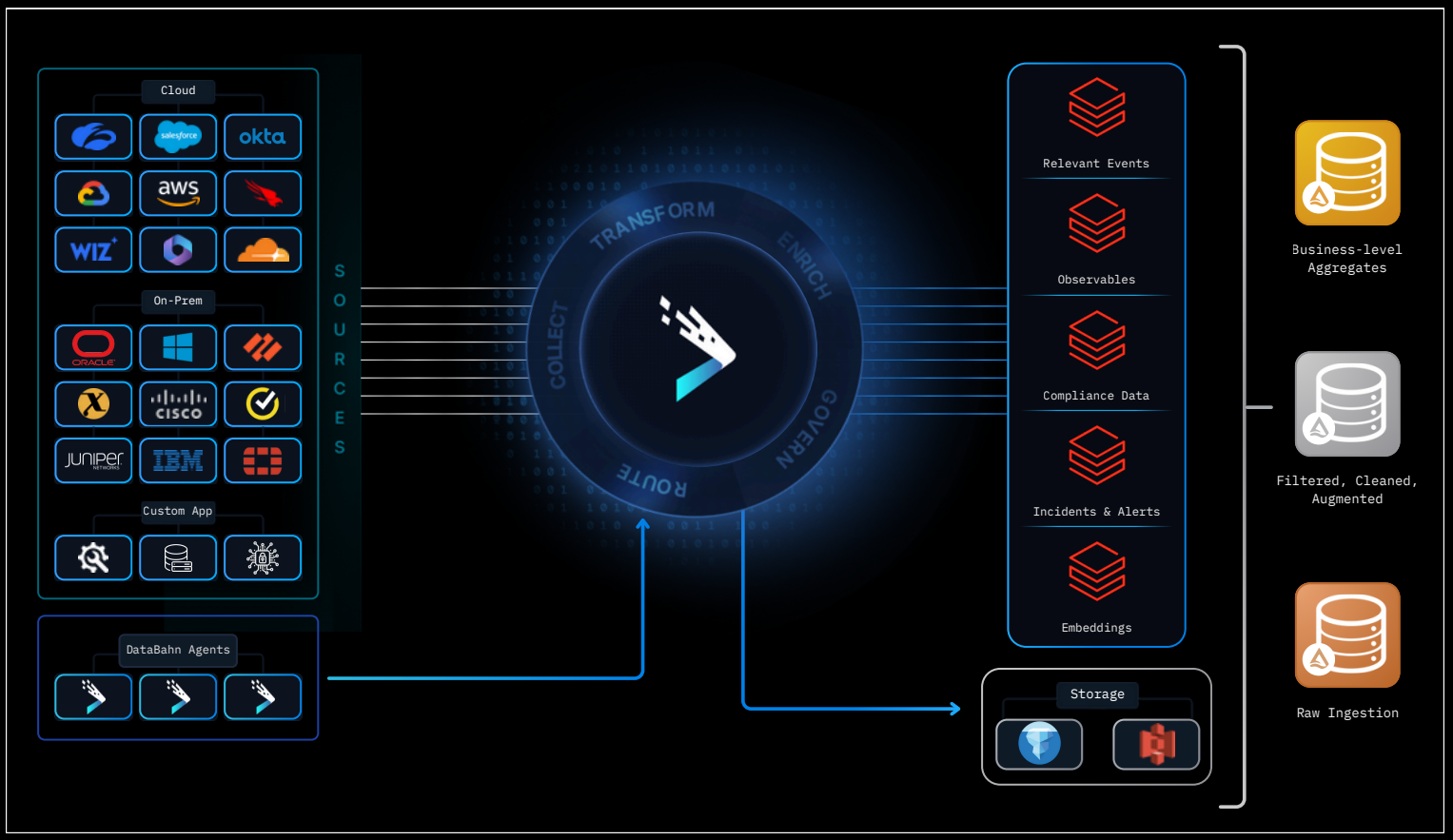
For years, enterprises have poured every sensor ping, workload log, and application heartbeat into SIEMs and data lakes under the assumption that more data meant stronger defense. Instead, this has produced terabytes of noise—repetitive, irrelevant, and fragmented across formats and tools. Analysts are left sifting haystacks while adversaries weaponize AI to move faster.

What’s needed is not just scale, but intelligence: the ability to prioritize and act on data as it moves. Databricks delivers the scale and flexibility to unify massive telemetry, while DataBahn adds intelligent collection, enrichment, and AI-powered tiering—turning raw data into trusted, actionable insight.

Making the vision real for enterprises

Security leaders don’t need more dashboards or security tools; they need their data to enable their teams to move faster and with more confidence. Their data needs to be reliable, contextual, and usable to enable faster and more effective work across tasks such as threat hunting, compliance, or powering a new generation of AI-powered workflows and systems.

By combining Databricks’ unified platform with DataBahn’s Agentic AI pipeline, enterprises can:



Simplify Ingestion into Databricks

- Use DataBahn's **500+** plug-and-play integrations and connectors with a wide array of products and devices.
- Automate the aggregating of data from any source – including custom applications and microservices – with AI-powered parsing and normalization of logs, metrics, events, traces, and even transactional data.

2. Convert logs into insights

- Filter out low-value telemetry before it clogs your Databricks lakehouse dashboards and tables, preserving what matters for detection and investigation. This reduces both the volume and the overall time for the queries to execute.
- Leverage AI-driven analytics and insights via **Reef** for data in flight to enhance data discovery, simplify reporting, accelerate querying via natural language, and get cross-domain and context-rich responses and insights.

3. Enhance Medallion Architecture

- Enable a structured data journey through bronze, silver, and gold tiers, ensuring that raw data is efficiently cleansed and enriched, ultimately delivering high-quality datasets for analytics and AI.
- Integrate with Unity Catalog to manage data governance, track data lineage, and ensure compliance at every stage of the Medallion pipeline.

- Standardize layer transitions to maintain data quality and schema consistency, while identifying and isolating data sets in transit to limit exposure.

4. Use best-of-breed services and technologies

- Leveraging DataBahn's simplified data orchestration capabilities, Databricks customers can use additional tools to implement a cyber mesh architecture without having to worry about locking your data within your vendor cloud.
- Using Databricks' marketplace applications with DataBahn forking out data streams to different tables within Databricks.

5. Get visibility into the health of telemetry generation

- By using the dynamic device inventory generated by DataBahn to keep track of devices to identify unexpected silences, log outages, and detecting any other upstream telemetry blind spots.

6. Reduce overall costs of operating Databricks

- Removing the need for any staging locations or custom integrations by taking advantage of DataBahn's native streaming integration to load data directly into tables.
- By routing less-frequently accessed data sets and keeping a copy of your logs using Data Highway to low-cost storage infrastructure such as your cloud storage (S3 / Blob / GCP storage) while adhering to the same data models and using Databricks external tables to access them.
- By adopting the use of open data formats like Iceberg and storing data older than your standard retention periods outside of Databricks and using Iceberg tables to access them

Why Security Teams Choose DataBahn

Plug-and-Play Integrations

DataBahn offers effortless integration and plug-and play connectivity with a wide array of **500+** products and devices, allowing SOCs to swiftly adapt to new data sources. Use AI to connect with and parse custom applications and microservices.

Context-Rich Enrichment

Enrich logs with threat intel, user, asset, and geo-data to boost the precision of detections and investigations.

Seamless Format + Schema Handling

Databahn supports seamless conversion into any data model of your choice, facilitating flexible and faster downstream onboarding in Databricks

Resilient Data Collection
DataBahn's highly resilient Smart Edge ensures that your team doesn't have to worry about single points of failure or managing occasional data volume bursts – the data collection never stops.

Sensitive Data Detection
Detect and isolate sensitive data in transit to strengthen compliance and reduce exposure risk.

Lower Costs, Higher ROI
Cut Sentinel costs by filtering out redundant and low-value logs using built-in volume reduction rules.

Relevance-based data orchestration
Tier and segment data based on relevance and move into different repos and tables so you can put purpose to your data.

With DataBahn and Databricks, unlock the power of your data and convert telemetry at enterprise scale into intelligence and insights. DataBahn's purpose-built data collection and orchestration platform enables your team to automate and optimize processes to deliver data into Databricks. This is the next era of security – and it starts with Data (bricks + Bahn). An AI-native foundation in which telemetry is self-optimized and stored in a way to make insights instantly accessible. Data is turned into intelligence, and intelligence is turned into action.

DataBahn.ai is a leader in AI-driven Data Pipeline Management and Data Fabric solutions, helping organizations transform their data operations with innovative engineering and advanced analytics technologies. With its Data Fabric platform and cutting-edge AI capabilities, DataBahn.ai is committed to empowering organizations to harness the full potential of their data for a smarter, more connected future.

Learn more at databahn.ai

DATABAHN