DataBahn for Snowflake

Building the Next-Gen, Cost-Effective, and High Performant Security Data Lake in Snowflake using DataBahn

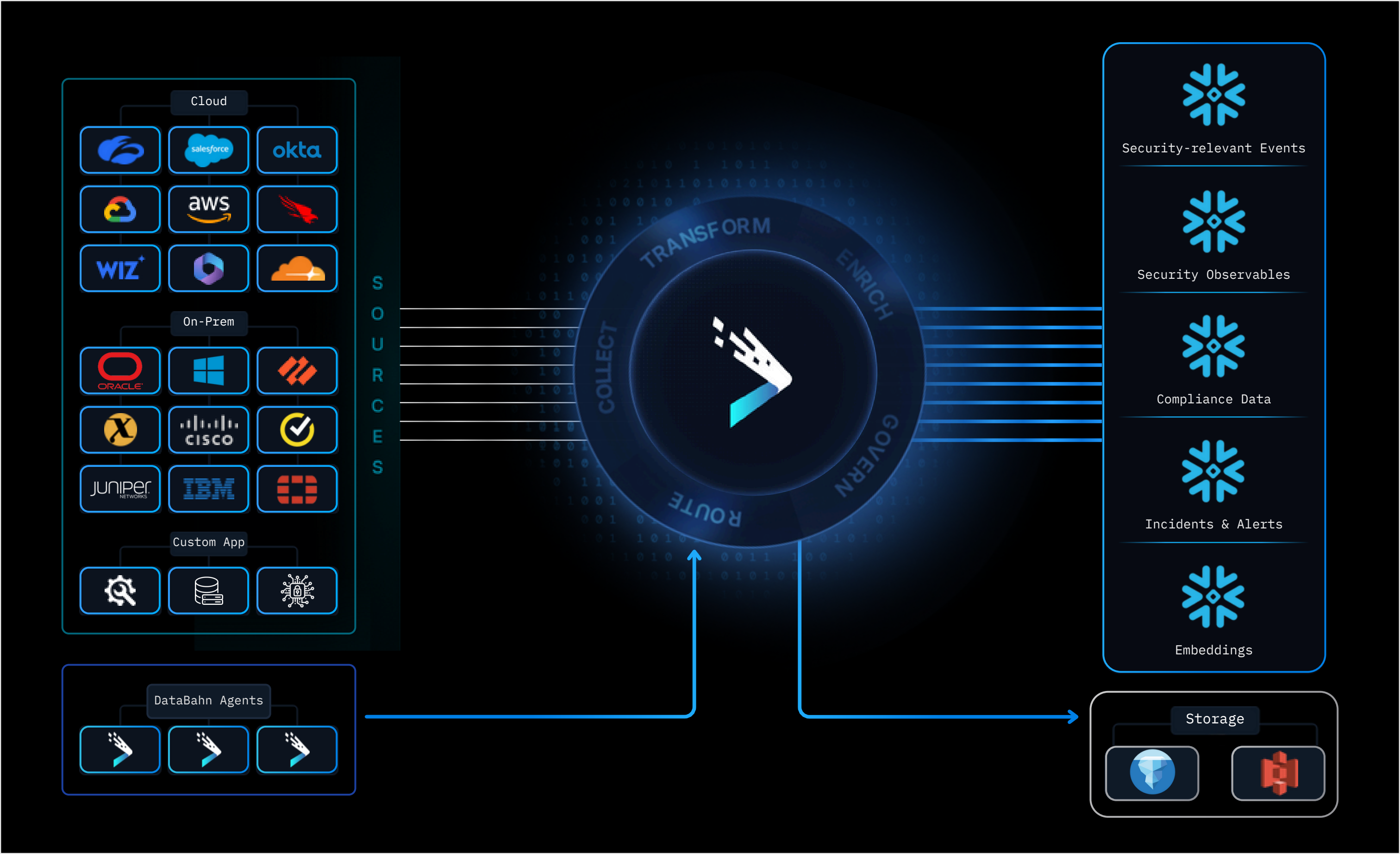
DataBahn + Snowflake

Snowflake is being chosen by many enterprises to be their data lake. Snowflake's schema flexibility permits the storage of raw data and on-the-fly schema application, ideal for the oft chaotic nature of data lakes. In performance terms, Snowflake's architecture excels in querying and analyzing large datasets, facilitating data lake analytics. Robust security features, encompassing role-based access control, encryption, and auditing, ensure data security and regulatory compliance. Data sharing is also straightforward, providing a secure means for collaboration and potential data monetization. With Snowflake's continued innovation and flexibility to bring in external tables and iceberg tables, it is becoming the default choice for many enterprises, especially because of its unique architecture which separates storage and compute. Snowflake has a lot to offer for security teams and is an excellent choice due to its ability to handle security workloads such as threat detections and ML-powered anomaly detection, threat hunting; but SOCs face challanges in managing custom data pipelines to centralize log ingestion, performance of queries, and unpredictable costs. This is why many security teams have been unable to realize the value of migrating from monolithic SIEMs to data lake powered SIEMs like Snowflake.

The Solution

DataBahn helps Snowflake users by streamlining data collection and ingestion and removing the burden of building customized integrations and customized pipelines, deploying staging locations, or managing your own Kafka clusters to take advantage of the recently-released Snowpipe streaming to take advantage of the near real-time data ingestion and availability within Snowflake.

DataBahn's purpose-built Smart Edge along with the Data Highway platform can take data from a range of sources (both on premise and cloud), parse and structure them into any format or data model of your choosing, enrich data with any meaningful context (internal and external), orchestrate the data to extract meaningful insights and deliver data and insights into Snowflake for optimal querying, high performant analytics and search, thereby reducing your overall operating costs with Snowflake.



Through DataBahn's Orchestration capabilities, SOCs and Security Teams can:

1. Simplify Ingestion into Snowflake

- Using DataBahn's plug-and-play integrations and connectors with a wide array of products and devices
- Using DataBahn's native streaming integration for a hassle-free, real-time data ingestion into Snowflake
- By effectively normalizing and structuring data using DataBahn's orchestration pipelines before the data is loaded into Snowflake tables

2. Convert logs into insights

- By using volume reduction functions like aggregation and suppression to convert noisy logs like network traffic / flow into manageable insights that can be loaded in Snowflake reducing both the volume and the overall time for queries to execute

3. Increase overall data governance and data quality by

- Identifying and isolating sensitive data set in transit thereby limiting exposure

4. Perform split-second threat hunting

- Use DataBahn's Indicator Index to extract insights such as Security Observables (IP Addresses, Domains, URLs, Hashes), Entity Relationships (Processes, Network execution, Registry Modifications), Intel context
- Using additionally derived context such as first observed / last observed time / frequency of observation to speed up data exploration and hunting

5. Use best-of-breed detection technologies

- Using Snowflake's marketplace applications with DataBahn forking out data streams to different tables within Snowflake

6. 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

7. Reduce overall costs of operating Snowflake

- By removing the need for any staging locations by taking advantage of DataBahn's native streaming integration to load data directly into tables
- By routing less frequently accessed data sets using Data Highway to low cost, cloud storage solutions such as S3 while adhering to the same data models and using Snowflake 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 Snowflake and using Iceberg tables to access them

Why Security Teams Choose DataBahn

Plug-and-Play Integrations

DataBahn offers effortless integration and plug-andplay connectivity with a wide array of products and devices, allowing SOCs to swiftly adapt to new data sources.

Enrichment against Multiple Contexts

Enrich data against various contexts including Threat Intelligence, User, Asset, and Geo-location, providing a contextualised view of the data for precise threat identification.

Always-On Data Collection

Smart Edge ensures uninterrupted collection, even during traffic spikes or outages, so your data never stops flowing.

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

Targeted Data Delivery
Route only high-value, security-relevant data to Sentinel, and offload the rest to Azure Blob or ADX, reducing cost while preserving access.

Seamless Format + Schema Handling
Auto-adapt to schema changes and format variations to ensure consistent data quality with minimal overhead.

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

With DataBahn and Snowflake, unlock the power of your data by maximizing the value while reducing the overhead it takes to collect and ingest data and the overall operating costs. DataBahn's purposebuilt data collection and orchestration platform enables your teams to worry less about data acquisition into Snowflake.

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