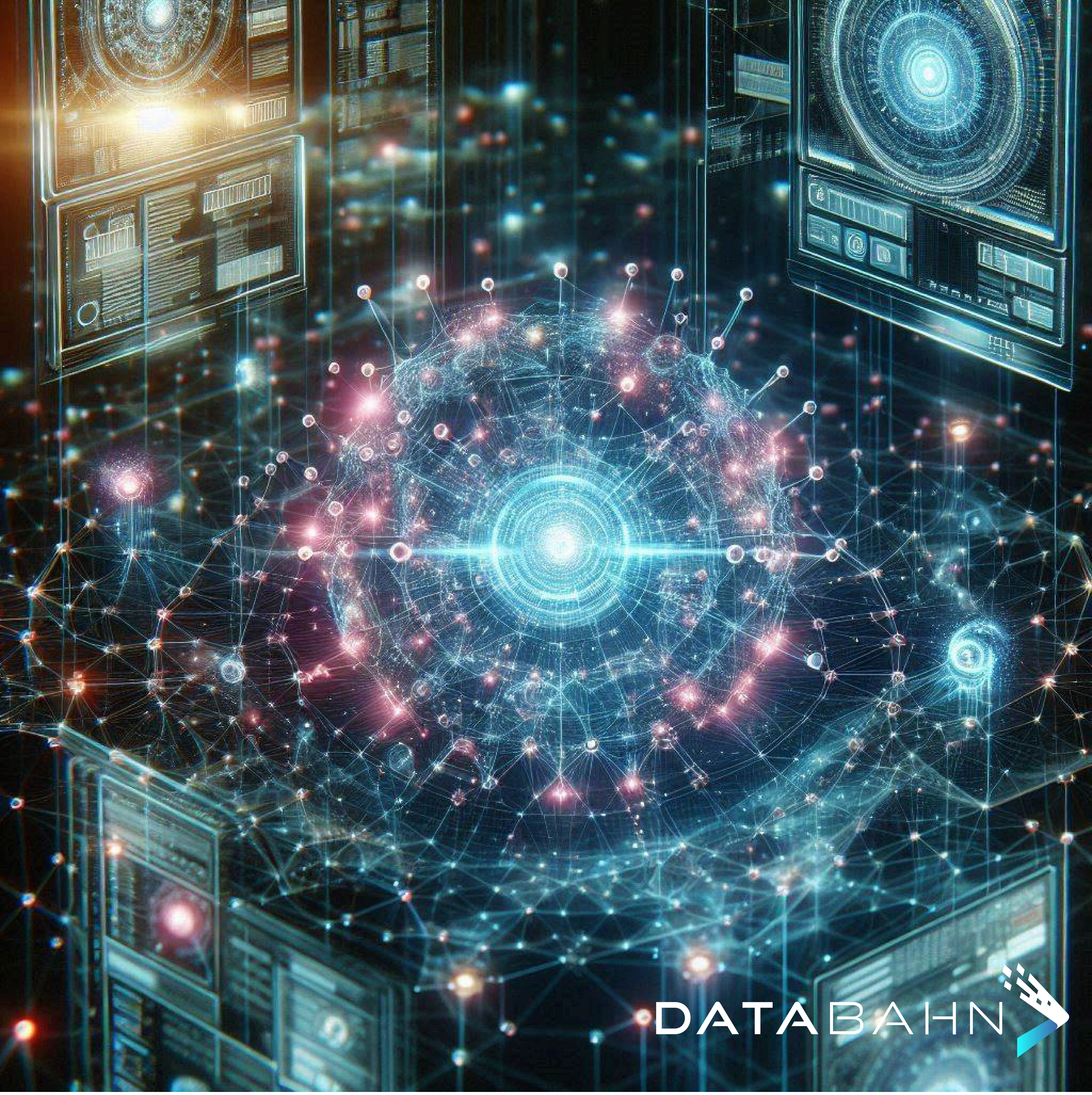


SOLUTION BRIEF

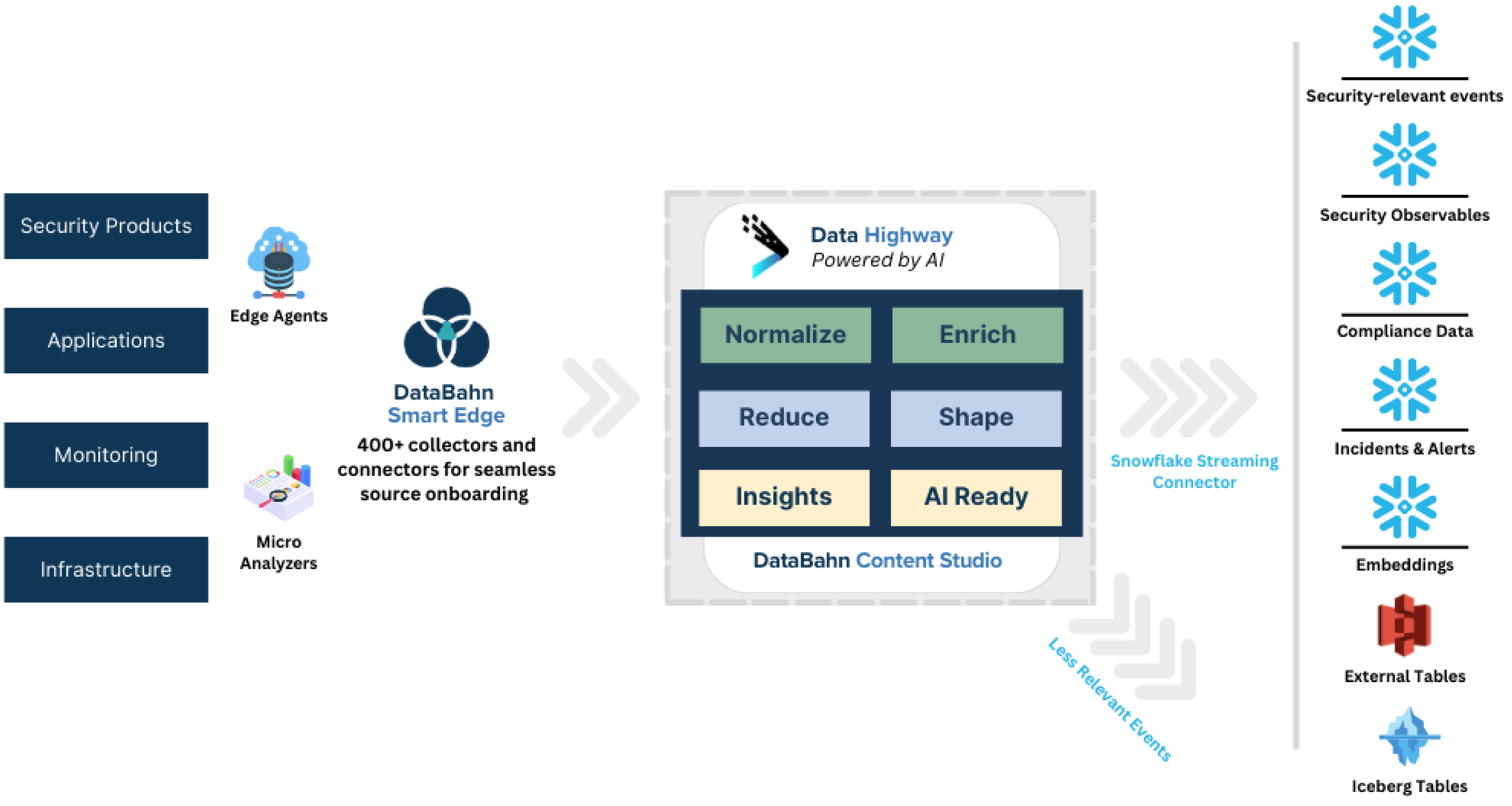
Building the Next-Gen, Cost-Effective, and High Performant Security Data Lake in Snowflake using DataBahn

snowflake

DataBahn for Security Data Orchestration in 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:

- **Simplify data collection and 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
- **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
- **Increase overall data governance and data quality by**
    - Identifying and isolating sensitive data set in transit thereby limiting exposure
- **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
- **Use best-of-breed detection technologies**
    - Using Snowflake’s marketplace applications with DataBahn forking out data streams to different tables within Snowflake
- **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
- **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

Benefits of using DataBahn

Out-of-the-box connectors and integrations

DataBahn offers effortless integration and plug-and-play 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.

Format Conversion and Schema Monitoring

The platform supports seamless conversion into any data model of your choice, facilitating flexible and faster downstream onboarding in Snowflake.

Resilient data collection

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

Reduced Costs

DataBahn helps you selectively extract key metadata based on frequency of usage, convert logs into insights to maximize retention of useful data whilst keeping costs of operating the warehouses optimal.

Risk-free data sharing

Use DataBahn to fork out data streams to different tables within Snowflake for restricted data sharing to Snowflake marketplace applications.

Get your data AI-ready

DataBahn's AI-ready framework gets your data cleansed, enriched, feature extracted, and with embeddings generated to build AI-powered apps.

Schema drift

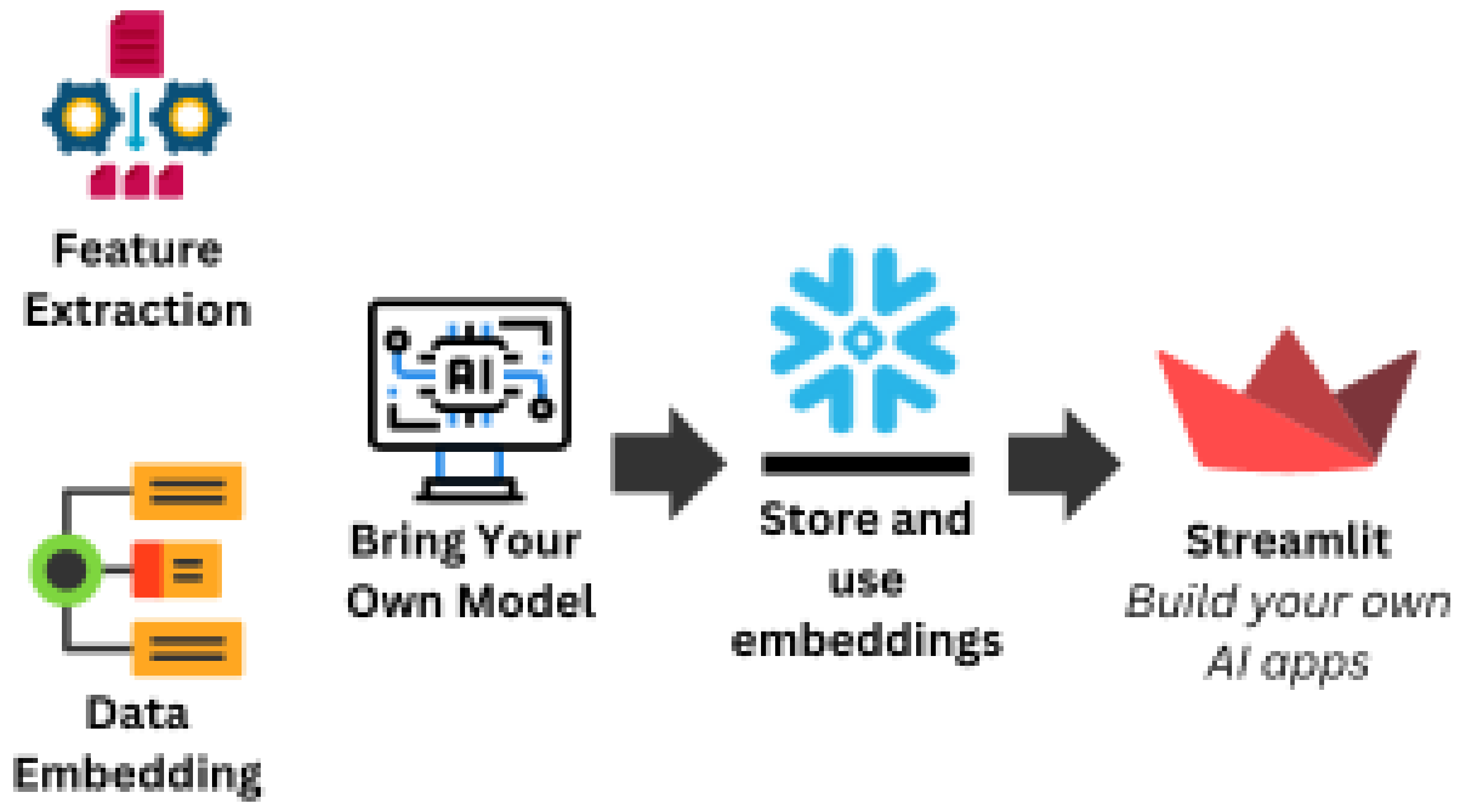
Detect changes to log schema intelligently for proactive adaptability and to avoid downstream detection impact.

Sensitive Data Detection

Identify, isolate, and mask sensitive data to ensure data security, governance, and compliance.

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.



Flexibility to your data stores

Deliver data to any external or iceberg tables and use Snowflake to centrally query and access the data.

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 purpose-built data collection and orchestration platform enables your teams to worry less about data acquisition into Snowflake.

ABOUT DATABAHN

DataBahn.ai's Data Fabric empowers organizations to optimize data management, reduce costs, and enhance security and IT operations. By integrating AI readiness, addressing IT observability challenges, and offering flexible solutions, the platform delivers significant operational efficiencies and strategic benefits, setting a new benchmark for cybersecurity data management in the digital age.

Learn more at **databahn.ai**

DATABAHN