# RESILIENCE CHALLENGES

## Making Data Collection & Orchestration Foolproof

DATABAHN

# INTRODUCTION

In the dynamic and challenging world of cybersecurity, security teams are constantly engaged in a relentless battle against complex threats while navigating through an overwhelming stream of data. Central to addressing these challenges are Security Engineering teams, whose mandate includes ensuring resilient practices in data collection, analytics, log management, and data storage. Despite the migration of various downstream platforms like SIEM, Security Data Lakes, and UEBA systems to the cloud, a substantial number of enterprises operate within a hybrid IT environment. This necessitates the transfer of data from on-premises infrastructure and cloud environments to the same downstream platforms.

In this landscape, many SIEM / Security Data Lake and UEBA platforms depend on remote agents and log collectors. These tools are pivotal in gathering logs and metrics, amalgamating them, and then transferring them to their cloud-based counterparts. However, the efficiency of these systems is often hindered by limitations such as single-threaded operations, the need for pre-planned capacity, and a dependency on manual intervention for adding capacity or rebalancing workloads. This whitepaper delves into these complexities and how DataBahn's **Smart Edge** and **Data Highway** products are engineered to ensure resilience in data collection, processing, and secure delivery, highlighting the critical role of effective data management in bolstering cybersecurity resilience.

Collection is typically one of the layers that SIEM or other log management platforms focus on outside of detection, storage, reporting, search, and compliance. In many organizations, this data collection layer has become the weakest link, especially when faced with unpredictable volume bursts from log sources like firewalls and proxies. Such scenarios often lead to back pressure in data pipelines and, in some cases, data loss.

Conventional data collection systems present challenges that hinder optimal resilience such as -

- Failure recovery
- Overhead on security engineering teams
- Impact on data distribution
- Time and resource drain

## THE AUTHOR


# Aditya Sundararam

***Chief Product Officer at DataBahn.ai***


Aditya is a seasoned product leader with an exceptional record of driving vision, strategy, and excellence in the technology startup landscape. He spent 10+ years as a product lead in Cyber Threat Analytics, harnessing a value-driven approach to product management and building innovative solutions. He has overseen global teams dedicated to developing and prioritizing SIEM and Security Analytics content. At DataBahn, he shapes and aligns our platform solution with market demands and customer needs.

## Failure Recovery

When failures occur during data collection, conventional platforms often lack efficient mechanisms for automatic recovery. This translates to added complexity as security engineering teams are burdened with manually identifying, addressing, and rectifying failures. The lack of automated recovery severely impacts the continuous and seamless flow of data.

## Overhead on Security Engineering Teams

Manual intervention for failure recovery imposes a considerable overhead on security engineering teams. They are compelled to divert their focus and resources towards identifying the root causes of failures, resolving them, and restarting the collection process. This shift in focus hampers their ability to proactively respond to security threats and incidents.

## Impact on Data Distribution

Similar challenges persist during data distribution to different destinations. Failures in this phase are exacerbated by the need to ensure the right data reaches the appropriate destinations. Without an automated recovery process, security engineering teams are faced with the task of managing the failed data delivery, causing further delays and disruptions. Capacity planning itself becomes a challenge as the bursts of traffic, frequently seen within the cybersecurity ecosystem, require dynamic scaling and optimal use of resources. Conventional planning creates single points of failure creating scenarios where resources on one side are over utilized while there are available resources on the other side that are underutilized.

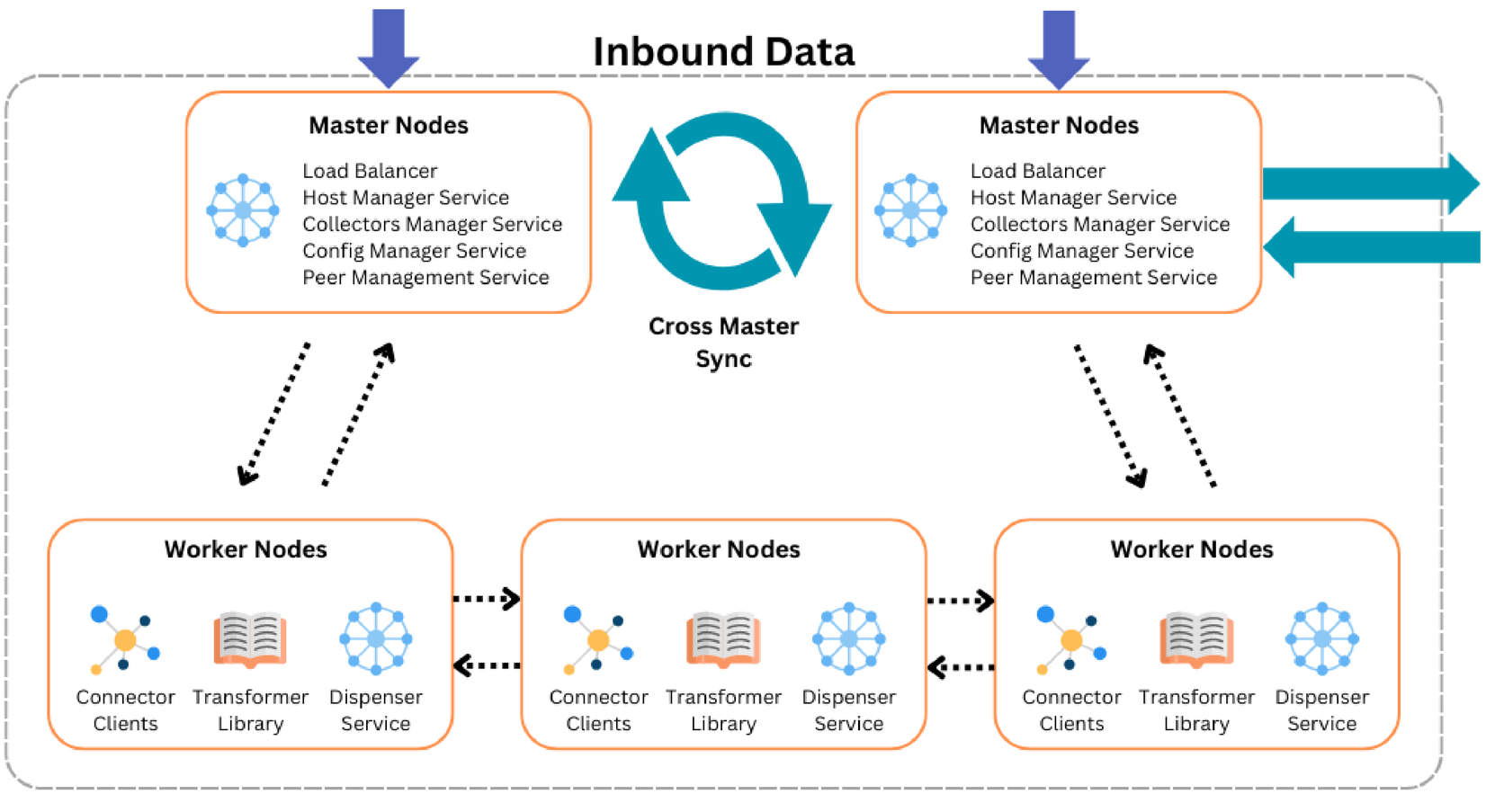
## Time and Resource Drain

The lack of automated recovery not only consumes valuable time but also exhausts resources within the security engineering teams. They are forced to invest excessive effort and manpower in identifying, diagnosing, and addressing each failure. This diversion of resources from proactive security measures to reactive recovery significantly impacts the overall security posture.

# The DataBahn Approach

## DataBahn Smart Edge: Empowering Data Collection and Processing

DataBahn's **Smart Edge** product is designed to excel where conventional agents and log collectors fall short. The DataBahn Smart Edge is a service mesh of log collectors, employing a leader-worker model for efficient data collection and processing.



## Purposeful Data Collection

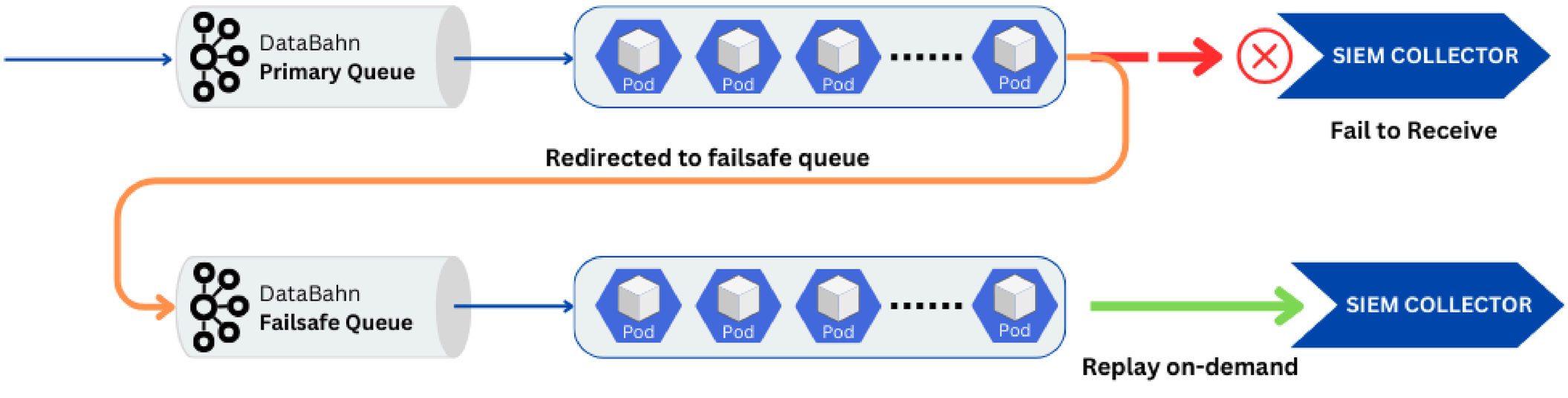
Purpose-built analyzers ensure that only pertinent data is identified and transmitted, optimizing resource usage, increasing data control, and reducing data volume to SIEMs or more expensive storage options.

## Auto Correction Mechanism

In case of service or node failure, the Smart Edge employs an auto-correction feature, guaranteeing uninterrupted data processing and continuous resilience without manual intervention.

# DataBahn Highway: Ensuring Resilient Data Delivery

The **Highway** product complements the Smart Edge solution, offering resilience during data delivery:




## Failover Handling

Understanding the criticality of every event within the security ecosystem, the Highway manages events that fail to be delivered due to destination limitations, ensuring resilient data delivery. There is clear data lineage maintained for auditability.

## Data Replay and Fail Queue

End users can choose to replay data later, ensuring no event is left unaccounted. The failsafe queue allows for efficient management and reprocessing of events. Conventional systems manage data in their buffer temporarily but start dropping events. DataBahn accounts for every event delivered.

## Data Lineage

DataBahn provides comprehensive data lineage, offering a clear and traceable path for each event, ensuring end-to-end visibility across the entire data lifecycle. Organizations can precisely track the journey of events, understanding their origin, processing steps, and transformations at a granular level. This facilitates effective auditing by enabling organizations to trace back to the source and identify any anomalies or issues at any point in the data flow. Data lineage in DataBahn enhances operational insight, allowing organizations to make informed decisions by knowing not just the current state of data, but its entire history and evolution through each processing step.

# CONCLUSION

DataBahn stands as a testament to resilience in data orchestration, directly addressing the challenges faced by security engineering teams. The Smart Edge and Highway products provide a seamless journey from data collection to delivery, ensuring optimal resource utilization, failover handling, and efficient data transformation. By streamlining these aspects and incorporating prepackaged security content, DataBahn sets a new benchmark in resilient data orchestration, revolutionizing the way security organizations secure and leverage their data.

# 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