

# Multi-Cloud Egress Cost Optimization

A Distributed Data Processing Approach for Modern Data Pipelines



## Executive Summary

In today's multi-cloud and multi-region environments, organizations generate massive volumes of telemetry across AWS, Azure, GCP, and on-premises systems. Whether operating across different cloud providers or multiple regions within the same cloud, centralizing this data for analytics and security operations often results in significant cloud egress costs, increased latency, and operational inefficiencies.

DataBahn addresses this challenge through a distributed, region-aware data processing architecture, enabling organizations to process data locally, reduce volume intelligently, and transfer only what is necessary.

By combining localized data processing, intelligent reduction techniques, and flexible routing, DataBahn helps organizations achieve 40–70%+ reduction in data transfer volume, directly translating into lower cross-cloud and cross-region egress costs, improved performance, and optimized data movement across globally distributed environments.



## The Challenge of Multi-Cloud Data Movement

Modern enterprises generate telemetry across multiple clouds—network devices, endpoints, SaaS platforms, cloud services. Traditional architectures centralize this data in a single cloud or SIEM before processing.

### Egress Cost Explosions

Cross-cloud transfers generate egress charges that scale with raw volume, not business value. Centralizing telemetry before processing creates unnecessary data movement and exponential cost growth.

### Latency & Scalability Limits

Funneling all telemetry through a central location creates latency and throughput limits. High-volume environments hit infrastructure ceilings, slowing data availability and analysis.

### Inefficient Data Movement

Raw, unfiltered data moves across clouds before any optimization. Noise, duplicates, and irrelevant fields consume the same bandwidth and cost as critical events.

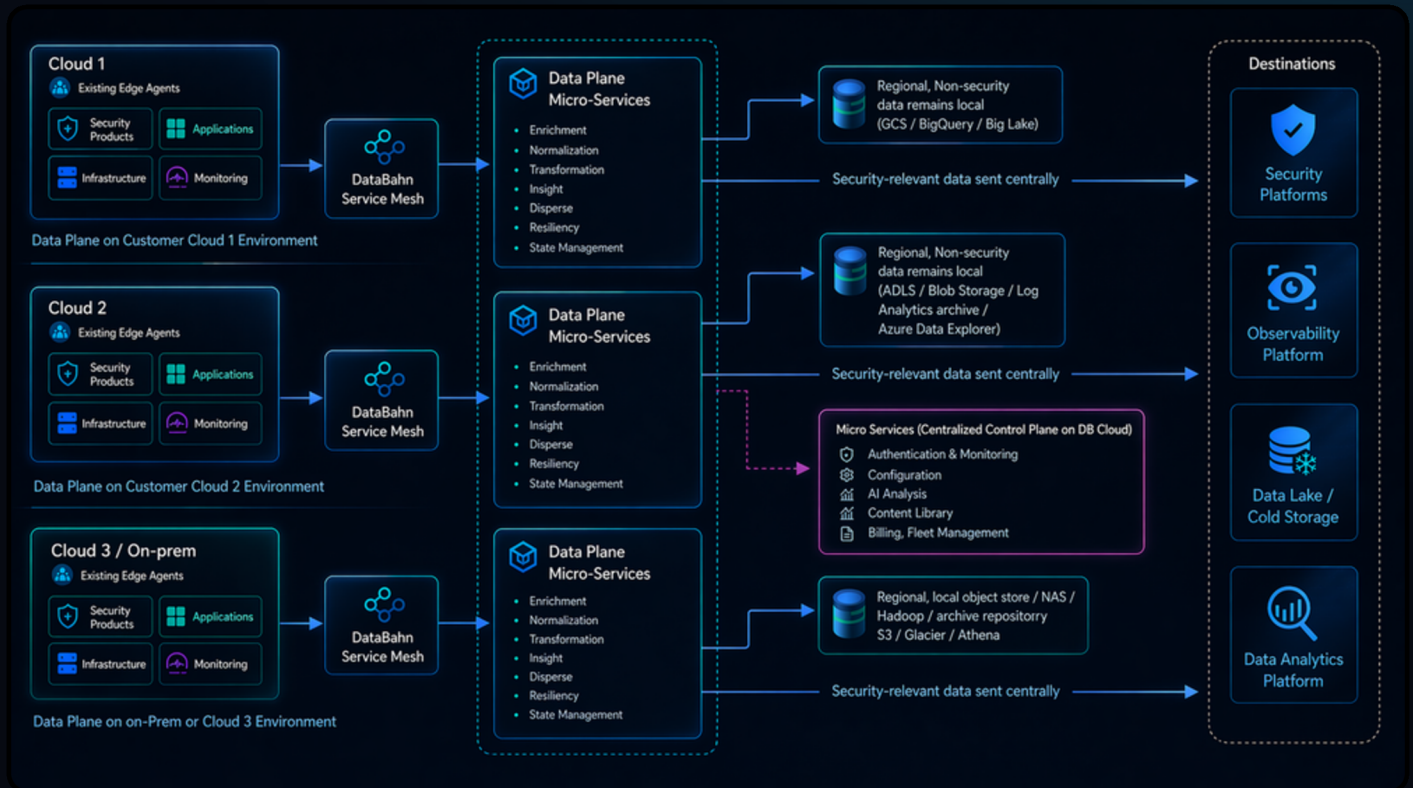
# DataBahn Approach: Distributed Data Plane Architecture

DataBahn introduces a modular, distributed architecture that separates:

- Data Ingestion Layer (Smart Edge / Collectors)
- Data Processing Layer (Data Plane)
- Control & Management Layer (Centralized Control Plane)

Core Principle: Process data where it is generated. Move only what is necessary. Access additional data on demand.

## Architecture Overview:



## Key Design Concepts:

**Local Data Planes** deployed within each cloud and region (AWS, Azure, GCP, On-Prem), enabling data processing close to the source

**Centralized Control Plane** for unified governance, monitoring, orchestration, and policy management.

**Intelligent Routing & Classification Layer** to dynamically separate and route security-relevant and non-security data

**Regional Data Localization Strategy** to retain non-security, low-value, and archival data within cloud-native regional storage (e.g., S3, GCS, ADLS, BigQuery, Blob Storage)

This architecture enables organizations to process data locally, retain non-security data regionally, and centralize only what matters, significantly reducing egress costs while maintaining full security visibility.

# How It Works:

## 1. Localized Data Processing

Instead of moving raw logs across environments, DataBahn performs **parsing, normalization, enrichment, deduplication, and filtering** within the same cloud or region where data is generated.

### Impact:

Reduces data volume before any cross-cloud transfer

Eliminates unnecessary movement of raw telemetry

## 2. Region-Aware Routing

For archival and analytics use cases, data is routed to local storage within the same cloud/region—avoiding cross-cloud transfers.

Examples:

- Azure logs → Azure Data Lake (same region)
- GCP logs → BigQuery / GCS
- AWS logs → S3 (same region)

### Impact:

Eliminates avoidable egress charges

Supports data residency and compliance requirements

## 3. Intelligent Data Reduction

DataBahn **aggregates repetitive events into summarized records, samples high-volume low-value logs, and forwards only required fields** to downstream systems—removing unnecessary attributes from raw logs.

### Impact:

Reduces total data volume transmitted across networks

Minimizes ingestion costs in downstream platforms (SIEM, analytics)

## 4. Multi-Destination Optimization

Single ingestion → multiple destinations (SIEM, Data Lake, Analytics) with different transformations per destination.

Example:

- **Splunk** receives filtered, enriched, security-relevant logs;
- **S3/Data Lake** receives compressed raw or summarized logs.

### Impact:

Avoids sending full raw data multiple times across clouds

Optimizes data format per destination use case

## 5. Compression at the Data Plane

DataBahn compresses data before transmission to downstream systems.

### Impact:

Reduces payload size during transfer—directly lowers egress bandwidth costs

Enables faster data transmission

## 6. Continuous Optimization via Insights

DataBahn Insights surfaces top talkers, high-volume sources, and event distribution patterns.

### Impact:

Enables data-driven reduction decisions

Continuously optimizes pipelines over time

DataBahn does not just reduce data volume—it redefines where and how data is processed, transformed, and routed, ensuring that only the right data moves across cloud boundaries.

This results in significant reduction in egress costs, SIEM licensing costs, and overall infrastructure spend, while maintaining full security visibility.

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](https://databahn.ai)

