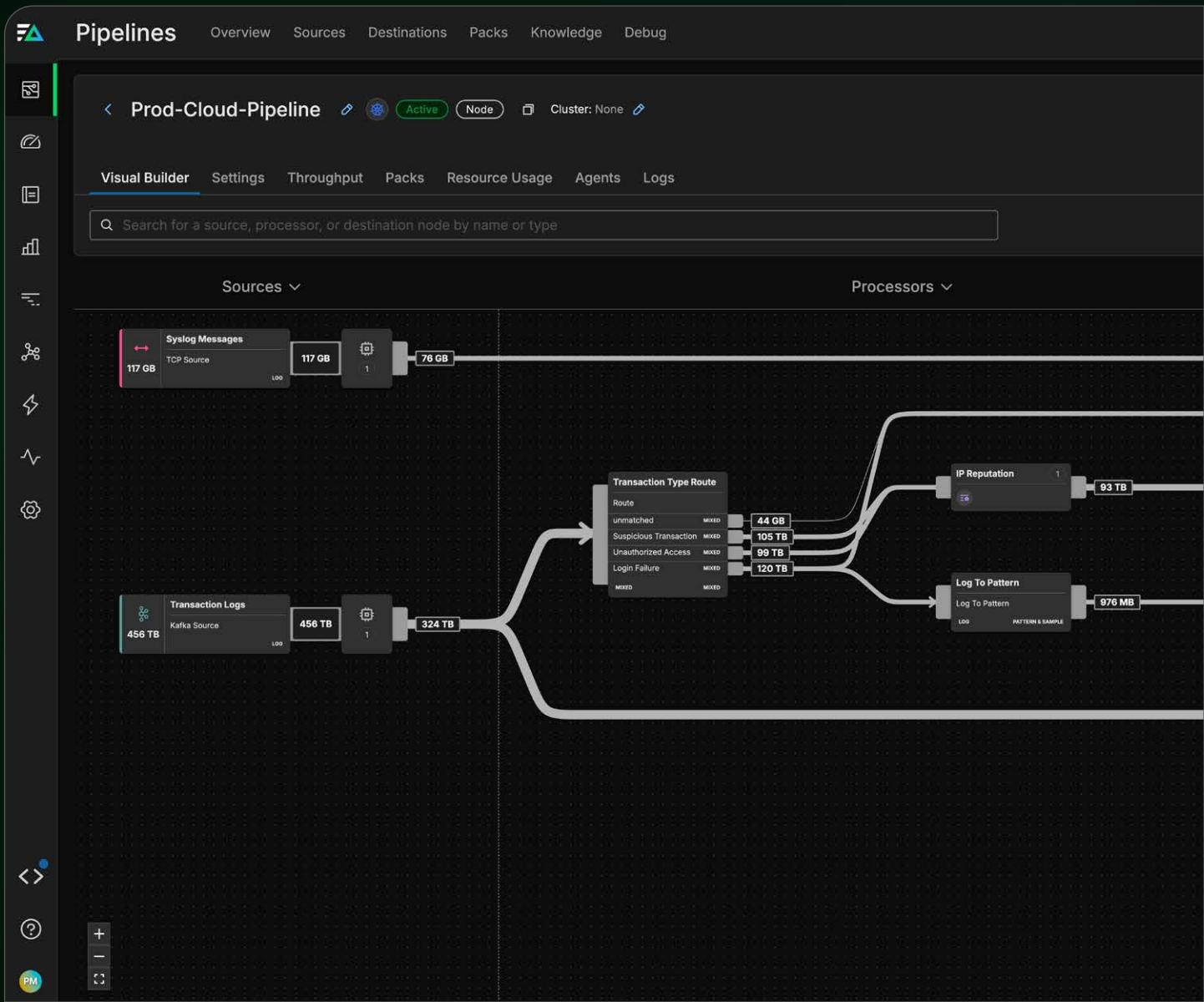EDGE DELTA

The Edge Delta Telemetry Pipeline Architecture

Optimization Rate

81.6%

335.5 TB

61.74 TB

Archives

Abstract

In this white paper, we discuss the Edge Delta architecture, a cutting–edge solution that forms the foundation for AI–native observability through intelligent data collection, processing, and routing. We cover the problems this architecture addresses — including accelerating telemetry data volume growth and heavy data source fragmentation — and explore the improvements it provides for AI readiness, performance, and cost efficiency.

The Problem

Accelerating Data Growth

Modern telemetry data volumes are increasing at a roughly 71% compound annual growth rate (CAGR). This is due in large part to the widespread adoption of microservice and cloud–native architectures, as well as the rapid proliferation of IoT devices:

71%
CAGR

Total Data Volumes

The need for next–gen architecture

High-Value Data

As data volumes skyrocket, organizations are forced to ship and store much more data in their observability platforms. This significantly increases observability costs and inundates downstream AI tools with large quantities of irrelevant data, negatively impacting MTTR.

Limitations of Traditional Pipeline Solutions

Organizations are increasingly adopting telemetry pipeline solutions to control and optimize their data more effectively. However, this can create significant issues if done incorrectly.

Performance

Several telemetry pipeline solutions on the market today claim to help enterprise organizations control and optimize data at scale. However, many of them are highly inefficient in practice and require additional VMs, hosts, or other infrastructure components to operate effectively. This places a significant burden on compute resources, driving up costs and undermining the performance benefits that telemetry pipelines are intended to deliver.

$$ Increased Overhead

Observability initiatives are highly complex and often require teams to integrate data from a variety of sources into cohesive, centralized workflows. Without a telemetry pipeline that can collect data from any source and in any format, standardize it, and enrich it with key metadata pre-index, teams are forced to manage a growing collection of brittle scripts or other custom-built solutions to do the job.

This is especially true for open-source solutions like the OpenTelemetry Collector and Fluent Bit. Configuring and maintaining these tools requires substantial time and resource investments, and they often lack features critical for enterprise environments — including robust performance, intuitive management interfaces, and comprehensive health monitoring.

AI Readiness

Application teams leveraging AI for enhanced insight and automated issue detection and resolution require clean, structured data to feed their foundational models. Without a telemetry pipeline that can normalize and enrich telemetry data before routing it downstream, these models lose much of their predictive and analytic value.

The Results

With growing data volumes, disparate formats and tooling, and a myriad of sources, modern data flows become incredibly chaotic, which makes it difficult to manage costs and enhance workflows with AI. With the wrong telemetry pipeline solution, teams not only fail to solve these problems — they also introduce entirely new ones around pipeline management.

The Edge Delta Telemetry Pipeline Architecture

Edge Delta's unified platform enables teams to optimize costs, streamline data management, and access intelligent insights without any tradeoffs. Let's explore the architecture that makes this possible.

Edge Delta's Telemetry Pipelines are powered by lightweight, highly-performant agents with built-in intelligence that can deploy in any environment, including Kubernetes clusters, Linux VMs, Windows Servers, IoT devices, and Edge Delta's hosted cloud. They are written in Go, a modern and efficient language, and provide support for multiple OS and processor architectures (including self-contained binaries for every platform). They are also fully configurable through Edge Delta's intuitive SaaS control plane, with local configuration options available for select environments.

These vendor-neutral agents enable teams to collect data from any source, optimize it pre-index with intelligent processing recommendations, and normalize it for downstream AI consumption. Additionally, their efficiency allows for advanced processing capabilities — including log-to-pattern conversion, tail-based sampling, and lookup-based enrichment — with minimal additional infrastructure.
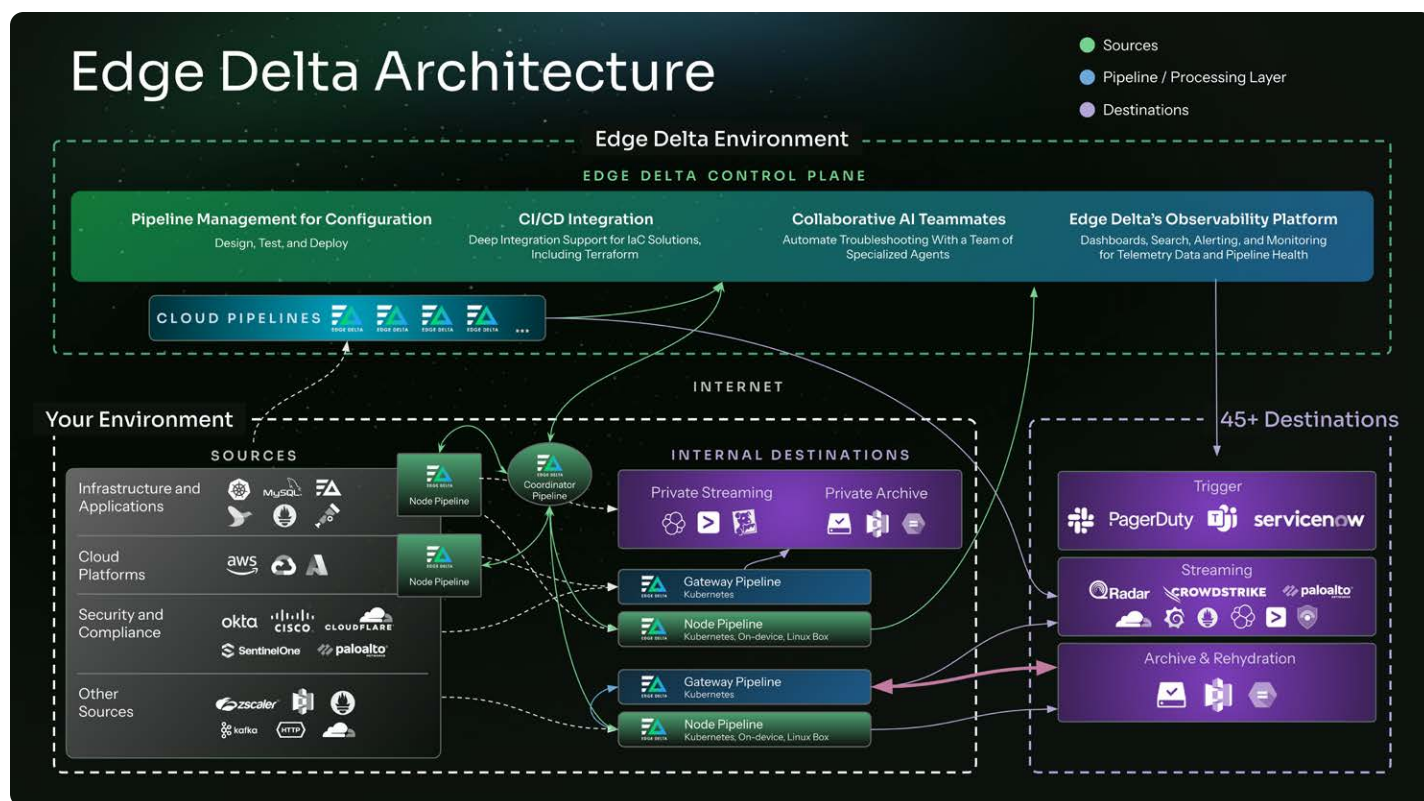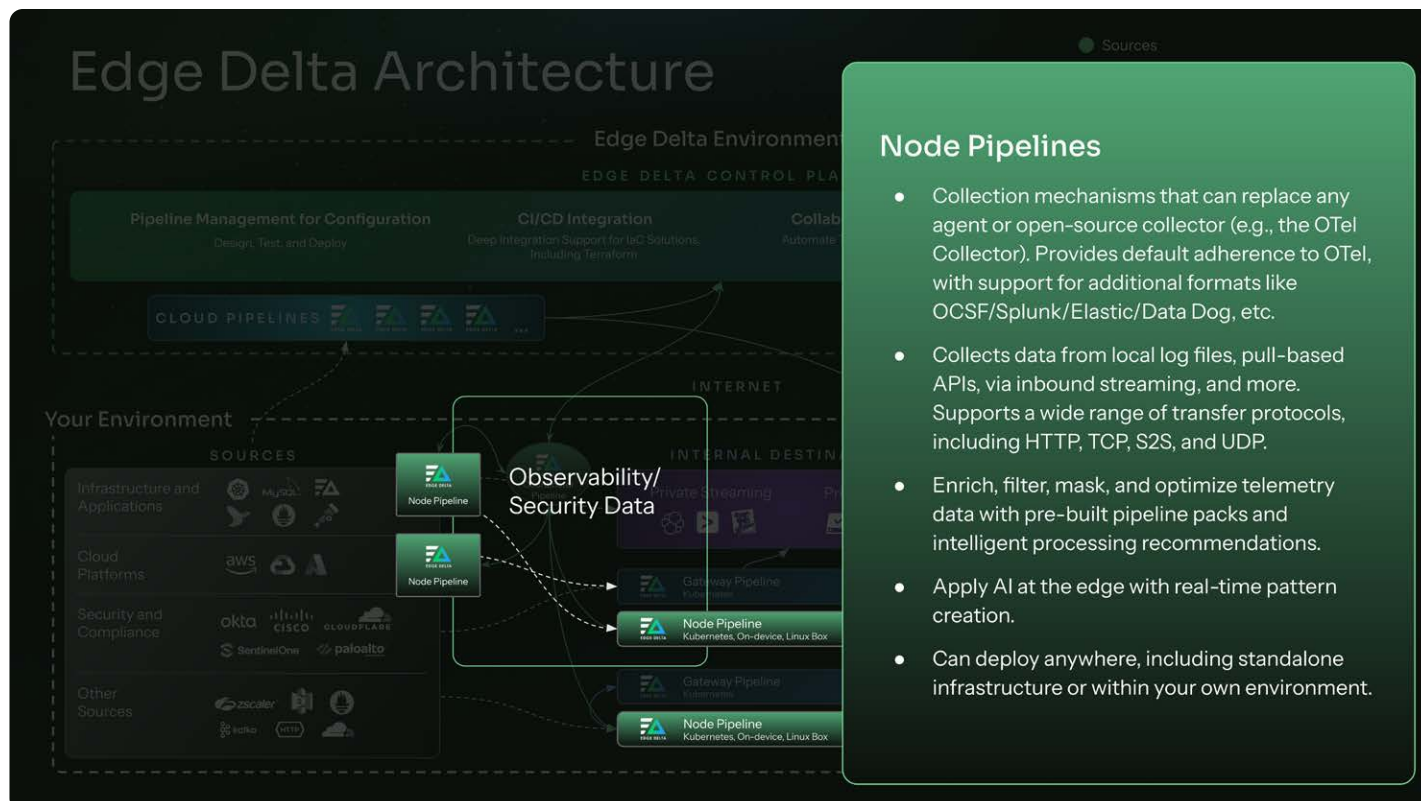

Edge Delta offers multiple pipeline deployment options suitable for different environments and infrastructure requirements. Let's go over each one.

Node Pipeline

In a Node Pipeline deployment, Edge Delta agents are installed directly within your computing environment, placing an agent on each host or compute source (for instance, on each node in a Kubernetes cluster as a daemonSet deployment).

By processing the data where it's created, these agents enable real–time data standardization, enrichment, filtering, and optimization at scale. They can extract critical insights, send alerts, and intelligently patternize data — all before routing it downstream to Edge Delta's Observability Platform, third–party monitoring tools, or archival storage.
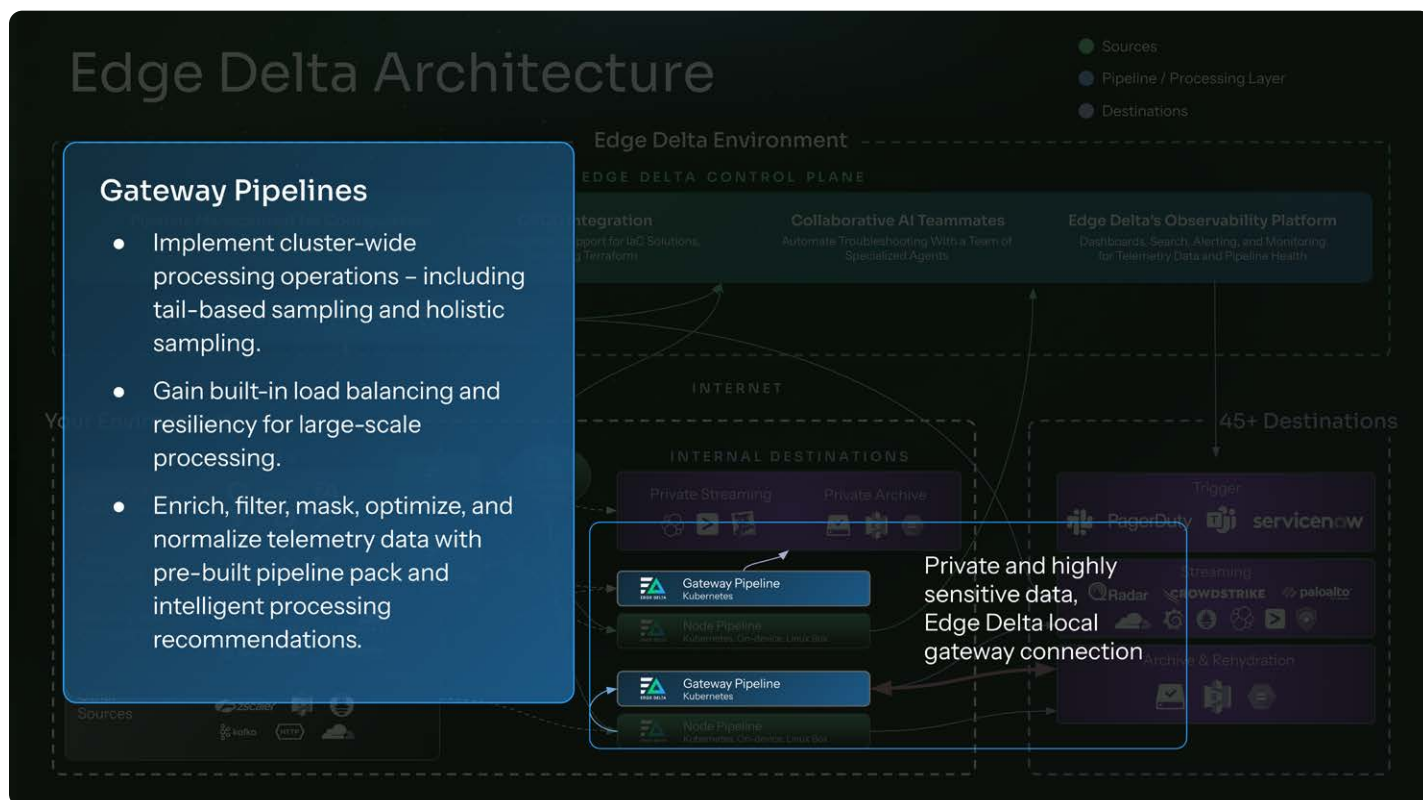

Node Pipelines

- Collection mechanisms that can replace any agent or open-source collector (e.g., the OTel Collector). Provides default adherence to OTel, with support for additional formats like OCSF/Splunk/Elastic/Data Dog, etc.

- Collects data from local log files, pull-based APIs, via inbound streaming, and more. Supports a wide range of transfer protocols, including HTTP, TCP, S2S, and UDP.

- Enrich, filter, mask, and optimize telemetry data with pre-built pipeline packs and intelligent processing recommendations.

- Apply AI at the edge with real-time pattern creation.

- Can deploy anywhere, including standalone infrastructure or within your own environment.

Gateway Pipeline

The Gateway Pipeline is a Kubernetes–specific deployment that serves as a centralized, pre–index aggregation and processing layer.

Its agents are deployed as a replicaSet, ensuring robust scalability and consistent availability, and it can ingest telemetry data from multiple sources, including Node Pipelines and other external sources. Once collected, the Gateway Pipeline supports cluster–wide processing operations, including service–level metric aggregations, environment–wide log deduplication, and holistic sampling.

A key feature of the Gateway Pipeline is tail–based sampling. As individual agents observe portions of a distributed trace within their respective parts of the environment, they forward those spans to the gateway. Once all spans converge in the gateway, it reconstructs the full trace, allowing users to apply holistic sampling logic to filter their trace data. This gives users the flexibility to make more informed decisions about which traces to retain or discard based on complete context, rather than applying sampling decisions by evaluating spans in isolation.

Node–to–Gateway connectivity is facilitated through a specialized gateway input that receives data from multiple agents, and an output in Node Pipelines specifically configured to send data to the gateway.

Coordinator Pipeline

The Coordinator Pipeline deployment serves as a control plane agent, enabling communication and coordination amongst Node Pipeline agents within a Kubernetes cluster.

It is purpose–built to streamline agent orchestration, acting as an intermediary for control signals between the Edge Delta Observability Platform and the agents themselves. This approach significantly reduces communication overhead for configuration updates and health check–ins in large–scale environments.

The Coordinator Pipeline is composed of one agent, deployed as a single–replica Kubernetes Deployment. There is always exactly one coordinator agent per cluster to avoid duplicate data collection and management confusion. It leverages the Kubernetes API to automatically discover other pipeline agents in the cluster, helping to establish logical groupings that contribute to a unified view of the environment. The Coordinator Pipeline is also required to enable Live Capture in clusters with more than one node.
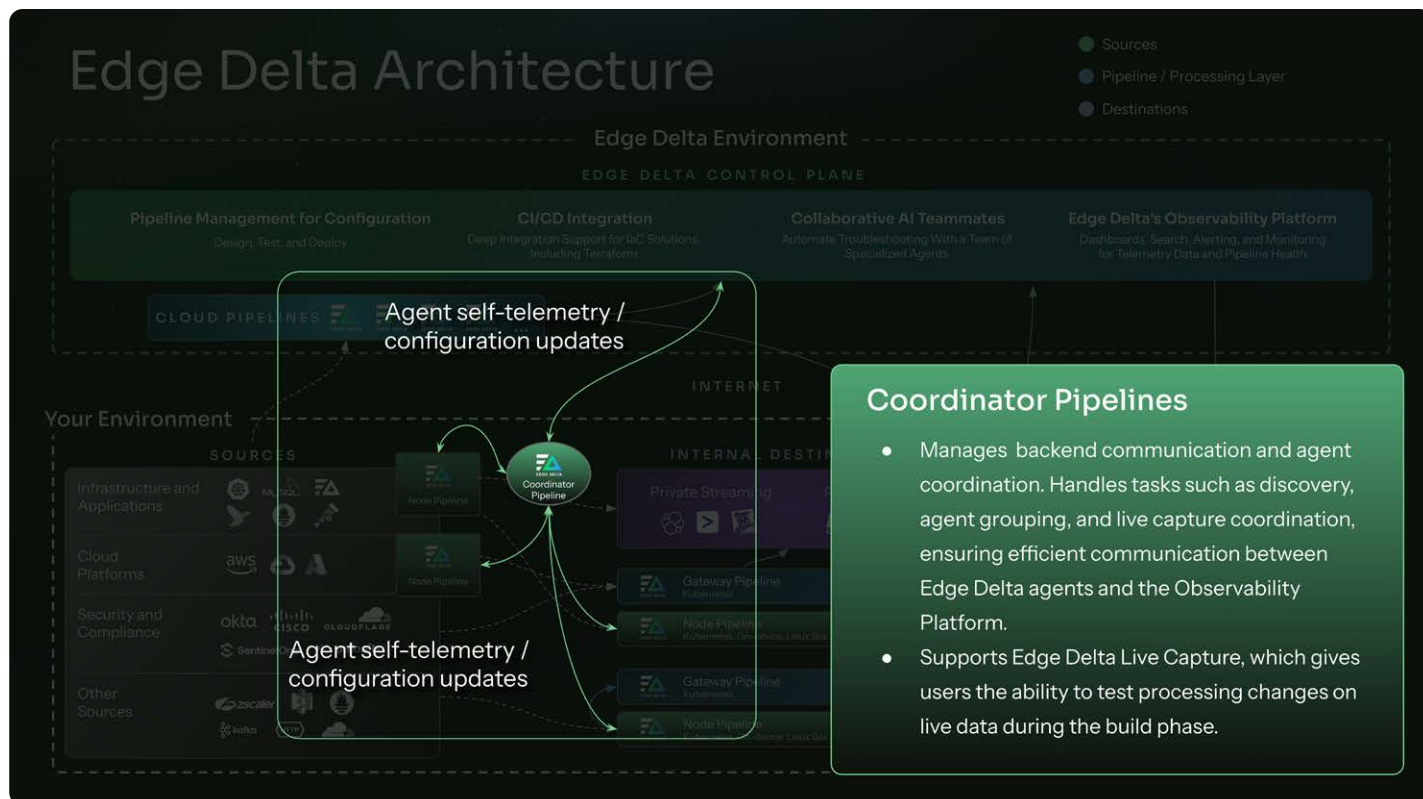
Cloud Pipeline

For teams that want the benefits of the Edge Delta architecture but don't have the ability to deploy new infrastructure, cloud pipelines offer a streamlined alternative. This model supports Node and Gateway deployments fully hosted in the Edge Delta cloud, delivering the same capabilities with minimal setup.
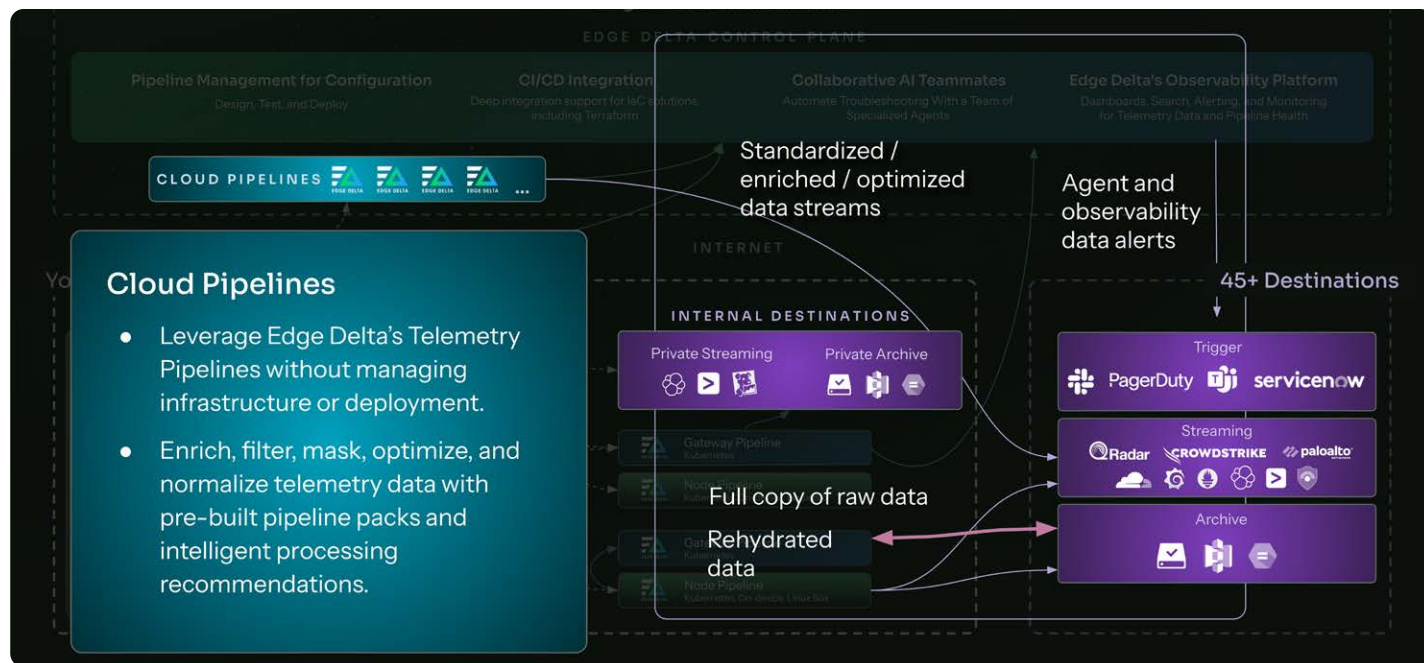
All deployments are managed through the Edge Delta Control Plane, which gives users:

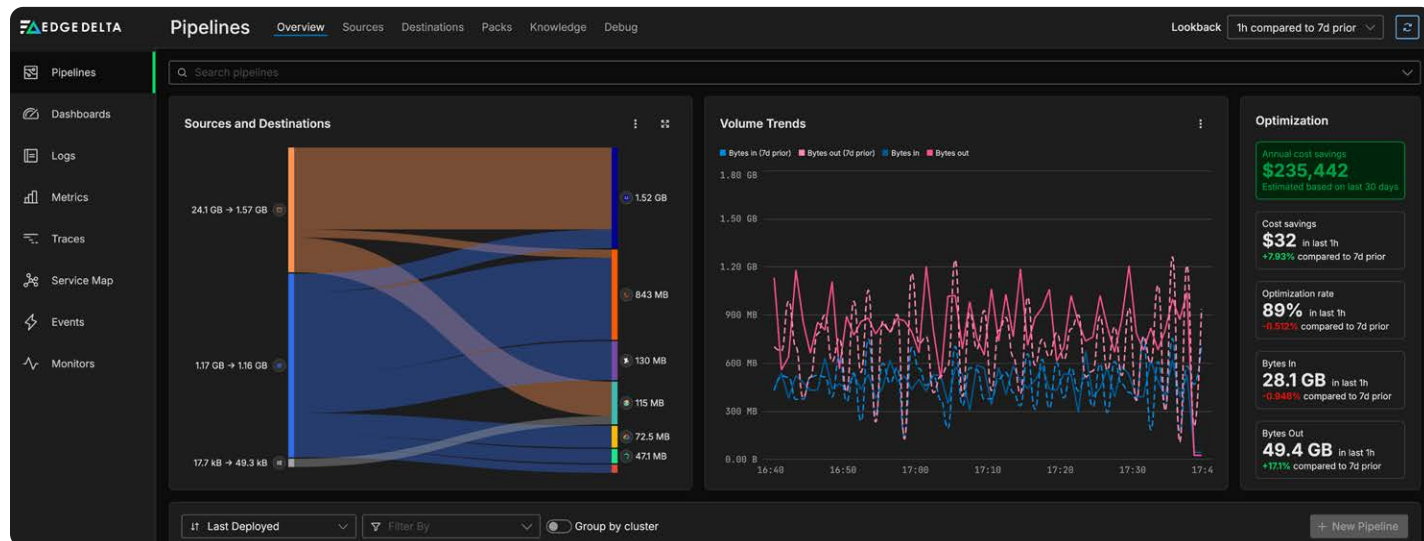| Deep observability for pipelines and data flows | An intuitive, graphical interface for designing, testing, and deploying pipelines | Rich CI/CD integration to automate configuration updates | Real-time alerting and intelligent patterns for all observability data |

Challenges Addressed

The Edge Delta Telemetry Pipeline architecture unlocks a number of key benefits for telemetry data management, particularly around infrastructure scaling, AI integration, and cost savings.
Let's walk through each one in depth.

AI–Powered Insights

Edge Delta Pipelines provide intelligent processing recommendations — including for masking, deduplication, enrichment, and filtering — while data is in flight. These recommendations are generated near–instantaneously for any data stream, including custom internal services and external sources with limited visibility, taking the guesswork out of data optimization.

AI–driven analysis also demands clean, structured, and high–quality data to deliver accurate results. Edge Delta pipelines make it easy to standardize data onto a consistent format, enrich metadata, correlate signals across sources, and remove irrelevant data, ensuring your downstream models are powered by the most robust and actionable datasets possible.

Horizontal and Vertical Scaling

Distributed workloads generate data at highly variable rates as demands scale up and down over time. As a result, observability teams need a data management solution that can scale dynamically — scaling up to meet spikes in demand and down to optimize cost.

Edge Delta's Telemetry Pipelines have built–in support for vertical and horizontal scaling, ensuring each workload receives exactly the resources it needs throughout its lifecycle. This approach minimizes costs while maintaining sufficient capacity to prevent failures.

Cost Savings

Edge Delta's pre–index processing approach often leads to a ~70% data optimization rate at enterprise scale. This results in massive cost savings, as low–value data isn't egressed from the source into expensive downstream destinations. The table below demonstrates the impact this optimization rate has on observability–related costs.

Costs calculated as follows: determine savings made if 30% of original volume is sent to the original observability platform with the remaining 70% forked to archival storage, using Edge Delta's Pipelines.

Data Volume	Original Observability Price (Per GB)	Edge Delta Pipeline Price (Per GB)	Amazon S3 Storage Cost (Per GB)	% Optimization Rate	Total Price	Price Reduction
100 GB	$2.00	$0.10	$0.03	70.00%	$72.10	$127.90
1 TB	$2.00	$0.10	$0.03	70.00%	$721.00	$1,279.00
100 TB	$2.00	$0.10	$0.03	70.00%	$72,100.00	$127,900.00
1 PB	$2.00	$0.10	$0.03	70.00%	$721,000.00	$1,279,000.00

Conclusion

The Edge Delta architecture addresses not only the core challenges of modern observability, but of legacy telemetry pipeline solutions, as well.

With real–time data processing, highly–scalable agents, and flexible collection and routing, Edge Delta's Telemetry Pipelines enable organizations to easily stay ahead of skyrocketing telemetry data volumes without compromising on visibility or budget. Edge Delta also delivers powerful AI enablement through pre–index data optimization, intelligent processing recommendations, and support for clean, structured data that drives more effective downstream analysis.

While development continues, our modern architecture delivers an enterprise–grade data management experience and establishes a scalable foundation for the future of AI–powered observability.