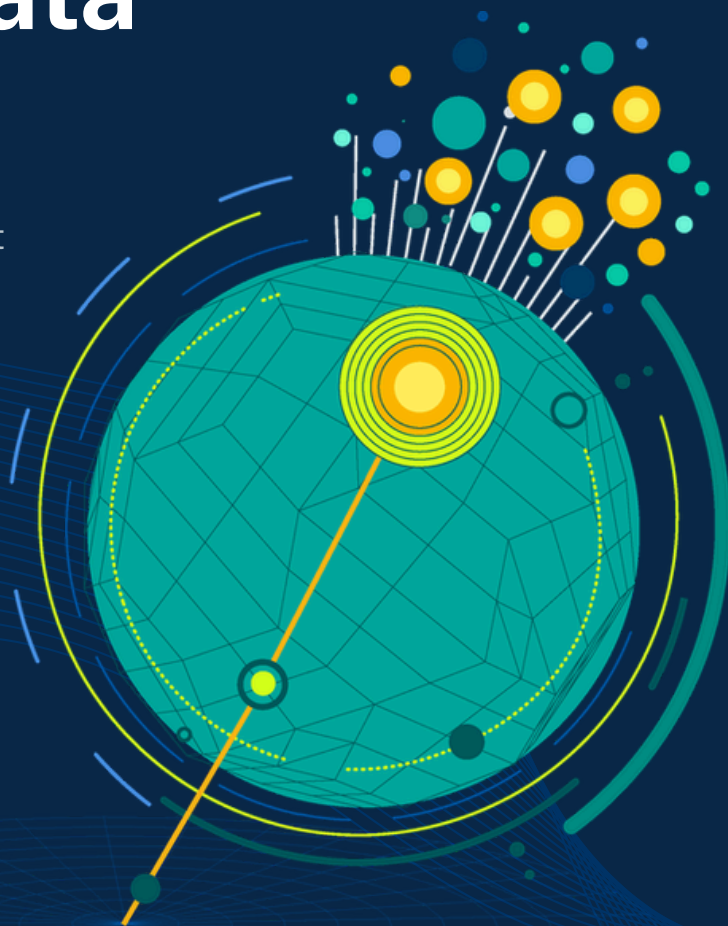


WHITEPAPER

Agentic AI Runs on Connected Data

How Hydrolix and TrackIt deliver a powerful, governed, AI-ready data platform for Media and Entertainment on AWS in under 10 days.



98%

ESTIMATED LOG STORAGE
COST REDUCTION

10

DAYS TO PRODUCTION

<2 min

TIME TO ROOT CAUSE

100%

DATA FIDELITY
RETAINED



TrackIt

Agentic AI Runs on Connected Data

Executive Summary	03
The Core Premise: Integration and Architecture First	03
The Problem: Why Disconnected Data Blocks AI Adoption	04
The Platform Architecture: Powered By AWS	05
Key Use Cases	06
Representative Scenario	07
Why This Solution Accelerates the AWS 'Powerful and Governed' Framework	08
Business Outcomes	09
Get Started with TrackIt and Hydrolix	09

Executive Summary

Every AI initiative in media and entertainment eventually runs into the same wall: disconnected data. Log data flows from multiple sources: Content Delivery Network (CDN) delivery, video players, ad insertion, rights management, and origin infrastructure, but it is scattered across tools and databases, and too expensive to retain at full fidelity. Agentic workflows require a unified, governed, AI-ready data foundation and a majority of organizations are restricted by siloed data.

The Hydrolix and TrackIt solution helps alleviate this issue by prioritizing deep platform integration and enterprise-grade data governance before any AI capability is layered on top. Connected data is delivered in under 10 days and is designed for organizations looking to build agentic AI workflows.

The Core Premise: Integration and Architecture First

Agentic AI is inclined to fail when data is disconnected or untrustworthy. Before any AI agent can reason across a media stack, three conditions must be met:

CONNECTED	Agents that cannot see all the data will produce partial or misleading answers. Every relevant data source must be connected via a unified data layer.
GOVERNED	Trustworthy AI requires trustworthy data. Consistent schemas, governance, and full fidelity data retention ensure agents reason over accurate, auditable information.
PERFORMANT	Interactive agentic workflows depend on queries that stay responsive as data grows into the petabytes. Architectures built on sampling or batch processing add latency between steps and pull agents away from the full picture they’re reasoning over.

The Problem: Why Disconnected Data Blocks AI Adoption

Before AI can reason across streaming operations, three compounding challenges must be addressed at the platform level.

01 Traditional Cost Structures Force Teams to Throw Out Data

A mid-size streaming platform generating 10TB/day can spend six-figure monthly to store that data in Datadog or Splunk. As costs rise, teams sample or drop the data that anomaly detection requires. Sampled data is a governance failure with direct AI consequences. An agent reasoning on incomplete data will produce incomplete and inaccurate responses.

02 Cross-System Correlation Is Manual and Slow

Diagnosing a viewer black screen means pulling from CDN, player, ad insertion, and encoder systems manually across multiple query languages. This integration gap is what prevents agentic AI from delivering meaningful results. AI agents cannot correlate data that remains trapped in disconnected systems, leaving every AI initiative constrained by the same data silos that limit analysts today.

03 Data and AI Are Not Accessible to The Right People

Quality of Experience (QoE) metrics, delivery health, and ad performance data often remain locked inside engineering tools, requiring specialists to retrieve and interpret them. Product managers, content ops, and executives cannot interact with the data directly. AI succeeds when a governed platform and its data can be securely accessed by the right people with the right interface.

WHY ARCHITECTURE MUST COME FIRST

Every agentic capability — whether automated workflows, anomaly detection, or natural language reporting — is only as effective as the connected, governed, high-performance data platform beneath it.

Hydrolix and TrackIt provide that platform by unifying data across the media stack while making it trustworthy, performant, and accessible. AI capabilities are not the starting point. They are the outcome of getting the architecture right.

The Platform Architecture: Powered By AWS

The solution is a three-layer platform: a universal data layer that connects every source, a managed services layer that delivers and operates the environment, and a natural language interface that makes the data accessible to every stakeholder.

Hydrolix, The Universal AI-Ready Data Layer

Hydrolix ingests time-stamped structured log data from virtually any system in the media stack. AWS-native sources include Amazon CloudFront delivery logs, AWS Elemental MediaLive encoder telemetry, and AWS Elemental MediaTailor ad insertion logs. Third-party sources include CMCD player telemetry, QoE platforms such as Mux, standalone ad servers, origin and encoder telemetry, rights management events, and Demand-Side Platform (DSP)/Supply-Side Platform (SSP) impression logs. Any source that produces a timestamped structured log becomes part of a single queryable layer. This creates the complete data foundation an AI agent needs to reason without blind spots.

Patented compression reduces storage costs up to 98% versus legacy tools, enabling full-fidelity retention for months or years, ensuring trustworthy anomaly detection and AI grounding. Sub-second queries on petabyte-scale datasets provide the performance baseline for real-time agentic workflows. Data flows via Amazon Kinesis and is stored in the customer's Amazon S3 bucket, with Hydrolix ingesting and managing the index, schema, and query layer. All data stays within the customer's AWS account, satisfying data residency requirements.

TrackIt, Managed Service and Implementation Partner

TrackIt delivers a production-ready environment in under 10 days using pre-built integration patterns for AWS-native and third-party sources. At go-live, TrackIt manages the Hydrolix cluster provisioning, connects log pipelines, builds and validates cross-system schemas, deploys the MCP server layer, and configures dashboards and alerting.

Support extends beyond implementation. TrackIt remains the operational owner, continuously monitoring pipeline health, onboarding new data sources as the environment evolves, and refining detection rules to maintain performance and accuracy.

The MCP Server: Natural Language Over Governed Data

The Hydrolix MCP server connects AI tools to the Hydrolix cluster enabling users to query data in plain English and get answers that synthesize data across every connected data source. Questions such as, "Why are viewers in the Pacific Northwest buffering this afternoon?" become cross-system analyses spanning CDN logs, player telemetry, ad insertion logs, encoder telemetry, and QoE metrics, with answers delivered in seconds without requiring engineering support.

Cross-system reasoning is possible because every connected data source is normalized and governed within a unified data layer.

Key Use Cases

Agentic Automated Workflows

Operations teams define workflows that trigger based on specific data conditions rather than based on schedules. Examples include:

- Automatically opening support tickets when CDN edge error rates exceed a threshold
- Routing escalations based on whether the root cause originates in the CDN, MediaLive encoding, or ad insertion detecting ad insertion failures
- Generating shift handoff summaries that synthesize overnight stream health and client-side behavior without manual effort.

TrackIt designs, deploys, and continuously manages these workflows as part of its ongoing managed service.

Anomaly Detection Across the Full Stack

Hydrolix retains full-fidelity data for up to 18 months by default at a cost structure that makes lookback practical. The system compares current performance against historical baselines, detects deviations in buffering or bitrate that precede viewer churn, and identifies infrastructure leading indicators before they impact viewer experience. With all telemetry unified in a single governed data layer, anomaly detection spans encoder output, CDN delivery, ad insertion, and player behavior without requiring manual correlation across multiple systems.

Natural Language Reporting for Business Users

Business users query governed operational data directly in plain English from day one.

Example queries:

“Show me QoE for our top 10 titles this week, by device type and CDN.”

“Which markets had the highest rebuffering during last night’s live event?”

“Did MediaTailor insertion latency correlate with viewer drop-off during peak hours?”

Results are generated from complete, unsampled data across every connected source, eliminating the need for engineering support or manual analysis.

Representative Scenario

A streaming platform with 5 million monthly active viewers goes live with the managed service. Within 10 days, telemetry from Amazon CloudFront, MediaLive, MediaTailor, a third-party ad server, player Common Media Client Data (CMCD) telemetry, and application logs is unified in Hydrolix. Dashboards and natural language queries are available to operations and business teams.

Two days later, viewers begin reporting black screen breaks for a specific audience segment. An automated workflow queries every connected data source in seconds, ruling out the CDN and encoder output before identifying a timeout in the third-party ad insertion server affecting player initialization on a specific device type. The platform automatically generates a prioritized alert with a complete diagnostic summary and routes it to the ad operations team, allowing the incident to be resolved before it impacts a broader audience.

OUTCOME

<2 min

FIRST COMPLAINT TO
ROOT CAUSE

Time from first complaint to root cause: under two minutes.

Without a unified data platform, the same investigation would require analysts to manually correlate data across four separate systems, delaying resolution by hours.

The outcome isn't faster AI — it's faster resolution, enabled by a connected, governed, high-performance data foundation.

Why This Solution Accelerates the AWS 'Powerful and Governed' Framework

To accelerate the vision of making enterprise data AI-ready requires a platform that is both powerful enough for enterprise-scale data and governed enough to produce trustworthy outputs, connected across the technology stack, and accessible to every stakeholder. This solution delivers on each of those principles.

POWERFUL

Sub-second queries on petabytes of data. Full-fidelity retention at up to 98% lower cost than legacy tools. Source-agnostic data ingestion across AWS-native and third-party source systems.

GOVERNED

Data stays within the customer's AWS account. Data residency requirements are satisfied by the architecture itself. No sampling, no gaps, no incomplete baselines.

CONNECTED

Includes native integration with Amazon CloudFront, MediaLive, and MediaTailor, as well as pre-built connectors for the most common third-party media sources. Every timestamped log becomes part of the unified foundation.

ACCESSIBLE

Natural language querying enabled by the Hydrolix MCP server lets every stakeholder explore operational data from day one without engineering support.

Business Outcomes

OUTCOME	DETAIL
Faster Incident Resolution	Identify root cause in seconds instead of hours by correlating data across every operational system. Automatically route incidents to the right team with complete diagnostic context.
Reduced Viewer Impact	Detect anomalies earlier with long-term, full-fidelity telemetry, allowing teams to resolve issues before they affect large audiences.
Lower Infrastructure Cost	Reduce log storage costs by up to 98% versus Datadog and Splunk, making full-fidelity retention economically viable at scale.
Business User Access	Enable operations, product, and business teams to query governed operational data in plain natural language from day one, with no engineering or SQL expertise required.

Get Started with TrackIt and Hydrolix

TrackIt and Hydrolix are working with AWS to help media and entertainment, gaming, and AdTech organizations deploy an AI-ready operational data platform. The joint offer is production-ready, with a repeatable deployment methodology, pre-built integrations for the most common AWS-native and third-party data sources, and a committed 10-day go-live.

Reference architecture, deployment runbooks, and MCP server configuration guides are available to support AWS field teams.

- Contact TrackIt or Hydrolix to schedule a technical discovery session.
- Request reference architecture and deployment runbooks via your AWS field team.
- Go live in under 10 days with a connected, governed, AI-ready data platform.

GET STARTED

CLICK HERE



TrackIt