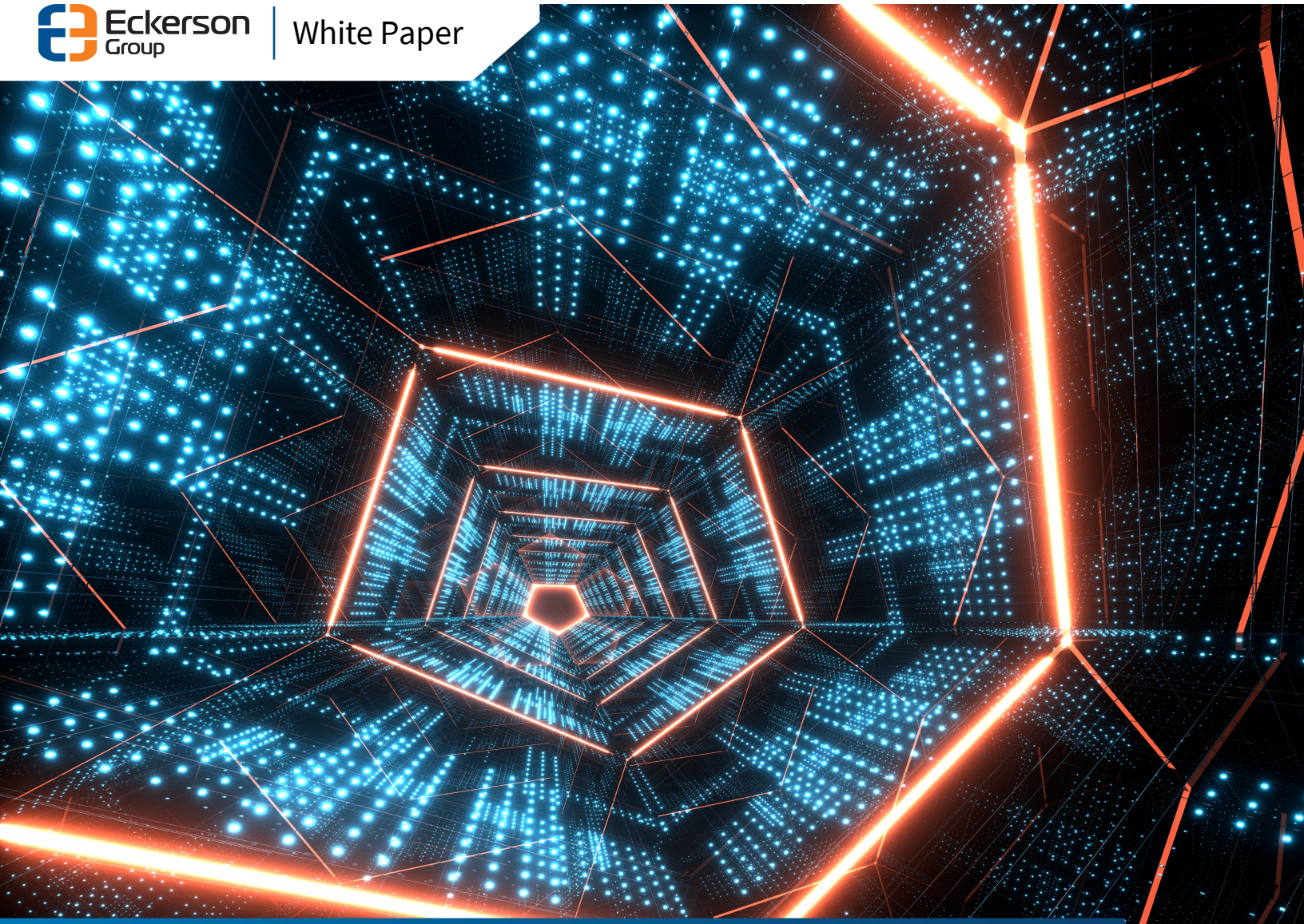
Eckerson Group | White Paper



Unravel Data

Data Observability for Modern Digital Enterprises

KEVIN PETRIE

JANUARY 2023

RESEARCH SPONSORED BY

unravel™

About the Author

Kevin Petrie is Vice President of Research at Eckerson Group, where he manages the research agenda and writes about topics such as data pipelines, cloud data platforms, machine learning, and data observability. For more than 25 years Kevin has deciphered what technology means to practitioners, as an industry analyst, author, instructor, marketer, and services leader. Kevin launched and led a profitable data services team for EMC Pivotal in the Americas and EMEA, and ran field training at the data integration software provider Attunity (now part of Qlik). A noted public speaker and author of two books on data streaming, Kevin frequently shares thought leadership content on **LinkedIn**.

About Eckerson Group

Eckerson Group is a global research, consulting, and advisory firm that helps organizations get more value from data. Our experts think critically, write clearly, and present persuasively about data analytics. They specialize in data strategy, data architecture, self-service analytics, master data management, data governance, and data science. Organizations rely on us to demystify data and analytics and develop business-driven strategies that harness the power of data. **Learn what Eckerson Group can do for you!**

About This Report

Eckerson Group provides independent and objective research on emerging technologies, techniques, and trends in the field. Periodically, we profile products that we believe provide exceptional value to customers. This report is sponsored by Unravel, who has exclusive permission to syndicate its content.

Table of Contents

Executive Summary

Unravel Data is a multi-faceted data observability platform that helps large, digitized, and data-driven enterprises automate the management of complex hybrid, multi-cloud environments. Unravel offers "DataOps observability" that addresses data application/pipeline performance, cost, and data quality at each stage of the iterative data lifecycle. Combining automation, full-stack observability, and AI-driven actionable intelligence, Unravel not only gives different members of data teams a "single source of truth" into what's going on and why, but also provides prescriptive recommendations on how to improve things. Unravel supports use cases for DataOps, performance management, financial operations, and planning and design. It differentiates itself with multi-modal functionality, automation, migration planning, self-service capabilities, and developer assistance.

Company

CEO Kunal Agarwal and CTO Shivnath Babu founded **Unravel Data** to automate DataOps tasks. They first focused on performance optimization and troubleshooting for Big Data pipelines that rely on data lake elements such as **Apache Spark**, Kafka, Hive, and Presto. Since then, the Unravel team has raised more than $107 million and grown to more than 100 employees. They also broadened their portfolio to address much more than data pipeline performance. Unravel offers "DataOps observability" that automates insights and prescriptive recommendations to improve not only the performance but also the cost of data pipelines and data/analytics applications, as well as the condition of the data itself throughout each stage of the iterative data lifecycle.

As part of this strategy, Unravel addresses multiple categories of observability with automation and AI-driven recommendations. This includes data pipeline observability, which spots and fixes performance issues with pipelines or underlying infrastructure elements such as processing engines and compute clusters. It also includes FinOps observability, which offers budget planning dashboards, chargeback reports, alerts, and recommendations on where and how to control cloud costs. In addition, Unravel now addresses data quality observability through integrations, starting with the open-source platform **Great Expectations**.

Target Customers and Use Cases

Unravel targets large, digitized, and data-driven enterprises that struggle to manage their business-critical data applications, infrastructure, and people in cloud-native, multi-cloud, and hybrid environments. They need to optimize performance on both heritage systems and new platforms while rationalizing how they consume cloud compute cycles. When performance does go sideways, their data teams must resolve issues fast to minimize business impact. They struggle to get the right visibility from legacy application performance tools that were not designed for modern data pipelines. Many of these enterprises, while cloud-first, also need help meeting performance and reliability SLAs on sprawling Hadoop data lakes that they first implemented a decade ago.

Enterprise data teams implement Unravel as a holistic observability platform to improve the stability and agility of their data- and analytics-driven applications. Unravel best serves enterprises that depend on cross-functional collaboration between data-related roles such as data engineers, data analysts, and data scientists; and business roles such as product owners, operations managers, and chief data or analytics officers. Enterprises such as **Adobe**, 84.51 (a **Kroger** company), and **Deutsche Bank** use Unravel to help data teams optimize data workloads across their stack. The primary use cases for Unravel span DataOps, performance management, FinOps, and planning and design.

- **DataOps.** Unravel provides monitoring, AI recommendations, and automated actions that help guide a DataOps program. It gives data engineers a granular understanding of their data environment as they build, test, and orchestrate data pipelines.
- **Performance management.** Unravel helps people developing data applications/pipelines manage performance themselves via self-service capabilities, reducing the burden on operations data engineers. Its AI and machine learning algorithms analyze millions of details in context to detect anomalies and identify performance problems, pinpoint root causes, predict risks, and automatically provide prescriptive recommendations.
- **Financial operations (FinOps).** Unravel also helps data teams and business teams alike address FinOps by automatically identifying opportunities for cost optimization and providing data-driven recommendations for improvement. For example, this might help run more workloads in the cloud with the same or fewer resources. Other capabilities include forecasting and allocating cloud resource costs to support budgeting and chargebacks by user, team, job, application, or infrastructure element. This helps cross-functional teams improve financial control and accountability for cloud usage.
- **Planning and design.** Armed with insights from the Unravel platform, data engineers and data architects can plan and design more effective infrastructure. They can provision and integrate data stores, containers, and compute clusters with a lower risk of bottlenecks. They also can predict the performance and cost implications of migrating workloads to the cloud. In-house developers, meanwhile, can use Unravel to build performance-friendly data applications.

Product Functionality

Unravel organizes its functionality into eight categories of entities and tasks, and offers each as a tab in its graphical interface. These categories are clusters, cost, jobs, pipelines, data, reports, automated actions, migration, and management.

- **Clusters.** The cluster tab provides detailed views of clusters and their resource consumption across the enterprise, allowing users to see their entire data estate. This can be broken down by workspaces, clusters, usergroups, and more. They provide support for all major modern data platforms across all cloud providers, including Databricks, Snowflake, Amazon EMR, BigQuery, Dataproc, and more.
- **Cost.** The cost tab helps users track costs to assist chargeback and budgeting by application, user group, resource, projects, or any user-defined tags that would allow an organization to capture business context.
- **Jobs/Pipelines.** Users study this tab to understand the status and characteristics of jobs or pipelines executed by elements such as **Spark**, streaming platforms, **BigQuery**, SQL, **Databricks** , Snowflake, and Kafka.

- **Data.** The data tab drills into the characteristics of tables and files to help understand data access, usage, and lineage. It also helps sort "hot" and "cold" data to support storage tiering.
- **Reports.** Users can generate reports on a variety of topics, including capacity forecasts, top resource-consuming jobs and users, tuning recommendations, and migration requirements.
- **Automated actions.** Unravel also lets users automate tasks that help resolve performance, cost, and quality issues. They can configure automated actions using templates or create custom ones, for example to stop rogue usage, improve performance of an app, or help a pipeline meet its SLA.
- **Migration.** The migration tab helps auto-discover and characterize target environments so users can anticipate migration requirements for workloads, datasets, and applications. It helps users explore cost and dependencies to plan a move to another environment.
- **Management.** Users handle a variety of administrative tasks in the management tab, for example, to configure monitoring, audit activities, and track billing for product usage.

Differentiators

Unravel differentiates its offering with multi-modal observability, automation, migration planning, self-service, and developer assistance.

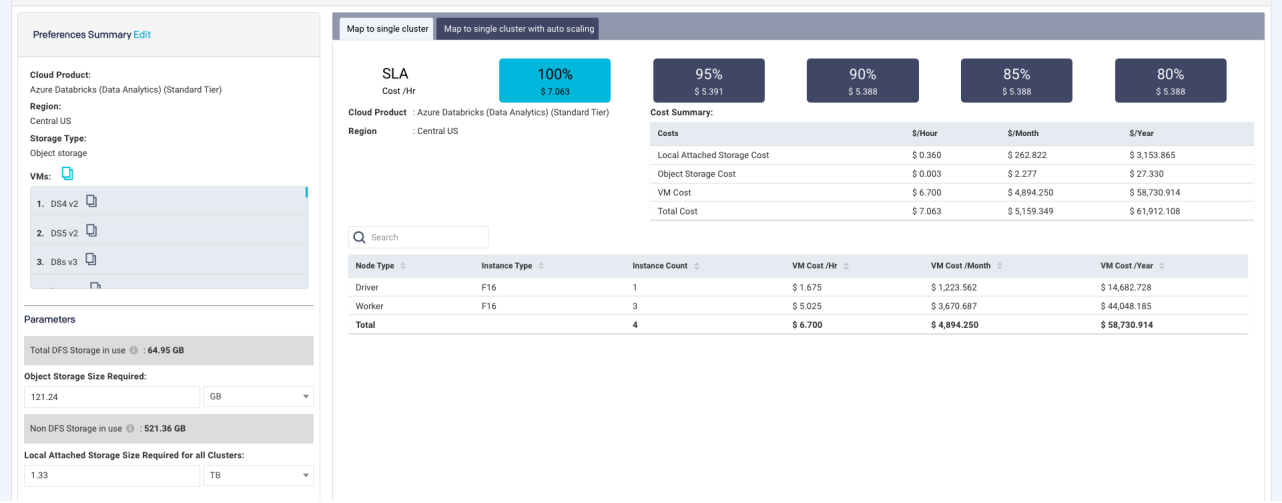
Multi-modal observability. Unravel enables data teams to monitor and optimize all the major dimensions of data delivery, including data pipeline performance, FinOps, and even data quality now with integrations into products such as Great Expectations. This multi-modal approach to data observability differentiates Unravel from point-focused performance tools, and complements rather than competes with data-quality tools. Unravel helps cross-functional teams collaborate as they troubleshoot current performance and improve their environments over time. They can use automated prompts and AI-driven recommendations to guide their actions as they address multi-modal observability.

Automation. Unravel reduces manual effort by automating tasks across the data lifecycle. For example, it automatically discovers issues before they happen, collects and correlates evidence, automatically pinpoints root causes, and recommends and acts to remediate issues on the user's behalf. In each of these cases data engineers, analysts, scientists, or operations teams play the role of overseer. They kick off automated tasks, receive the results, then review AI-driven recommendations about what to do next. The recommendations can get specific, for example, by prescribing the fastest or least-expensive way to run a pipeline and predicting the likely impact on KPIs such as compute cost or response times. Automated capabilities such as these dramatically reduce work for data engineers.

Migration planning. Unravel helps data teams plan migrations in three ways. First, they can model how workloads would behave in a target cloud environment to forecast risks, constraints, and expected costs. Second, they can automatically discover source cluster hardware, services, and utilization levels to scope their requirements for target environments. And third, they can compare source and target environments to identify any discrepancies in tools, platforms, or versions that need attention prior to the migration. Intelligence like this enables data teams to understand how best to host and optimize workloads in target environments before starting the migration.

In Figure 1, Unravel forecasts the likely costs of supporting a given SLA to varying degrees on a target cloud platform, in this case **Azure Databricks**, to assist migration planning.

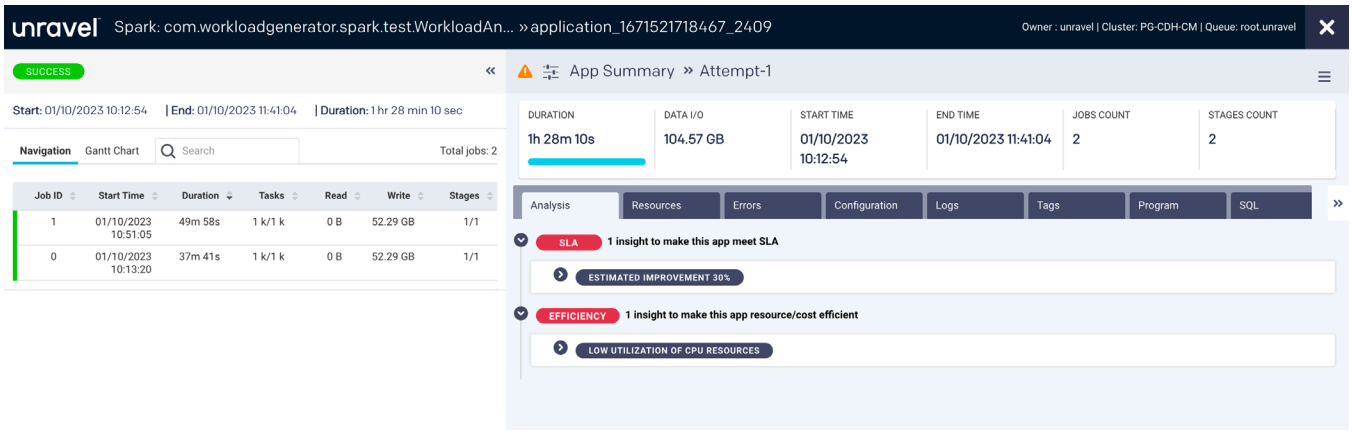
Figure 1. Compute Cost Forecast to Support an SLA



Self-service. Data engineers, data analysts, data scientists, and application developers can use Unravel to reduce performance risk themselves rather than enlisting the help of support or operations team members. Unravel prompts users to tune their own queries, for example, by flagging opportunities to rebalance data and workloads across pipeline resources. Users then can serve themselves by reviewing this AI-driven advice and approving it so Unravel can take automated action. It also empowers users to police themselves, for example, by notifying them of runaway jobs that threaten to inflate compute costs, crash systems, or throttle important applications. Users can then review and approve AI recommendations to cancel, reschedule, or re-configure such jobs before they cause issues.

In Figure 2, Unravel flags performance issues for an application and recommends ways to fix them. Users can click on each issue to get the details, then review and approve or deny the recommendation.

Figure 2. Performance Issues and Recommendations



Developer assistance. Data engineers and application developers integrate Unravel into their software development lifecycle, for example, to assist user acceptance testing and quality assurance. They can use Unravel to auto-profile the resource requirements of pipeline and application code before promoting that code to production. This information helps them design, build, and deploy higher-quality software—and more resilient architectures to support them. Developer features like these differentiate Unravel in the data observability space, underscoring the DataOps benefits of its platform.

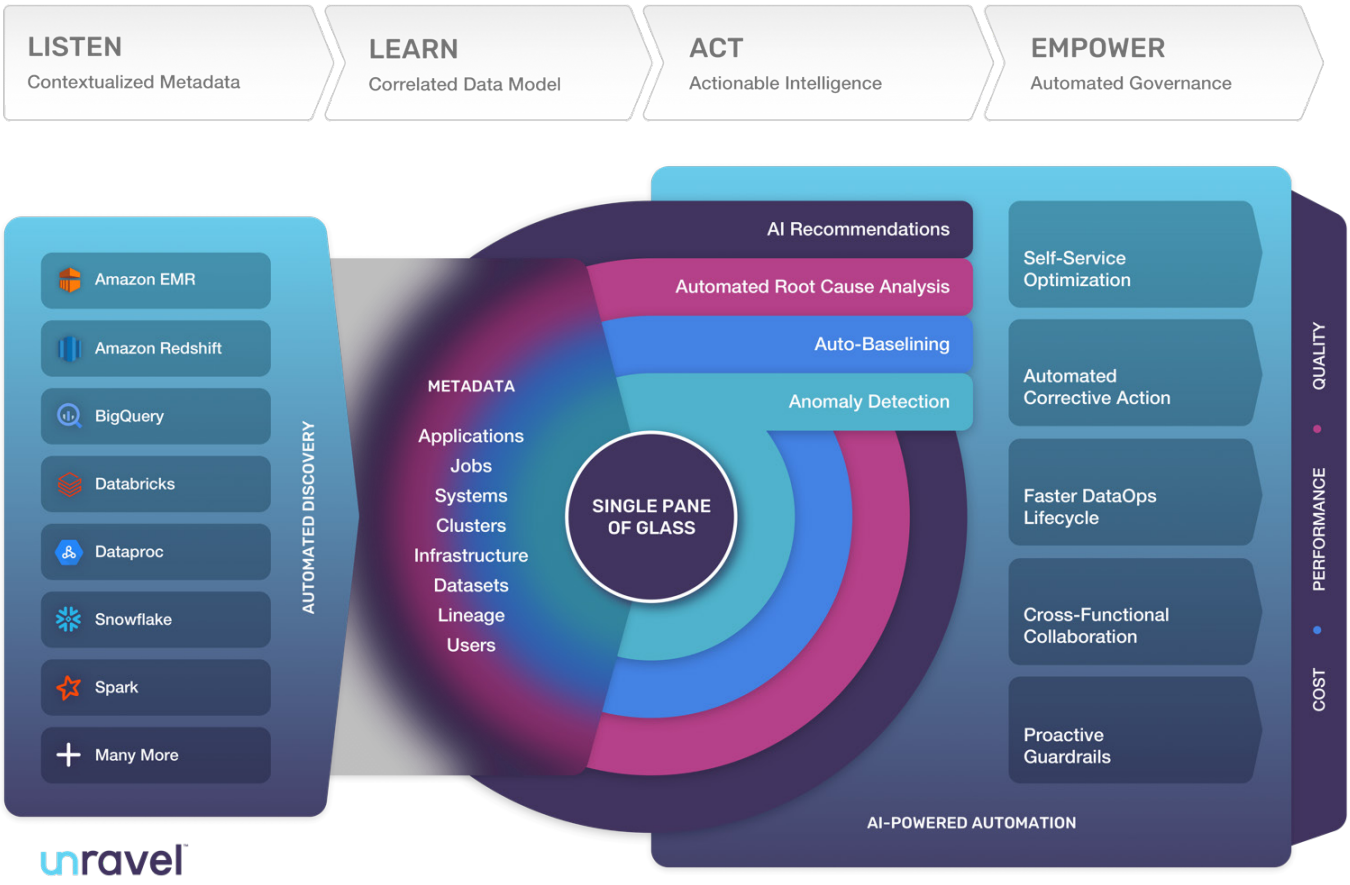
High Level Architecture

The Unravel product comprises telemetry data, **metadata**, a **correlated data model**, and **outputs**. Unravel integrates metadata from various architectural elements. These include data platforms such as **Snowflake**, **BigQuery**, and **Databricks**; the **Kafka** streaming platform; processing engines such as Spark; and compute instances such as **Amazon EC2**, **Azure Virtual Machine**, and **Google Compute Engine**. As these elements run jobs to support applications and data pipelines, they emit metadata such as logs, traces, configuration settings, events, and metrics. Unravel uses the metadata to build a correlated data model that analyzes how the various jobs and elements relate to one another. The model provides the logic for monitoring events, describing patterns, and identifying anomalies. This helps infer outcomes and take actions such as performance tuning.

Based on this model, Unravel generates outputs such as dashboards, reports, recommendations, and automated actions. It also sends alerts and notifications to issue tracking tools such as **ServiceNow**, **Jira**, and collaboration platforms such as **Slack** and **PagerDuty**. Unravel interoperates with the ecosystem of data pipeline elements, including ingestion tools such as **Fivetran**, orchestration tools such as **Apache Airflow**, transformation tools such as **dbt**, and BI tools such as **Looker**.

Figure 3 summarizes Unravel's platform capabilities for enterprise environments.

Figure 3. Unravel Platform Capabilities



Pricing

Users can start with a full-featured free version of Unravel for up to 50,000 units, which could be node hours/database units (DBUs) on **AWS** or **Databricks**. They can scale up to 100,000 node/DBUs for a fixed monthly fee, and receive training and onboarding assistance. For usage above this level, customers can obtain annual custom pricing with volume discounts. Unravel has no limits on the number of users or workspaces.

Summary and Recommendation

Since its founding, Unravel Data has expanded its portfolio to address data pipeline observability, FinOps observability, and now data quality observability through integrations with external tools. Data teams at large, digitized enterprises implement Unravel as a holistic observability platform to improve the efficiency, stability, and agility of their data/analytics applications. Unravel helps them address use cases such as DataOps, performance management, FinOps, and data quality while collaborating with business owners that consume data. Enterprises should consider Unravel if they seek a data observability product that offers multi-modal functionality, automation, migration planning, self-service, and developer assistance.

About Eckerson Group

Wayne Eckerson, a globally-known author, speaker, and consultant, formed **Eckerson Group** to help organizations get more value from data and analytics. His goal is to provide organizations with expert guidance during every step of their data and analytics journey.

Eckerson Group helps organizations in three ways:

- **Our thought leaders** publish practical, compelling content that keeps data analytics leaders abreast of the latest trends, techniques, and tools in the field.
- **Our consultants** listen carefully, think deeply, and craft tailored solutions that translate business requirements into compelling strategies and solutions.
- **Our advisors** provide one-on-one coaching and mentoring to data leaders and help software vendors develop go-to-market strategies.

Eckerson Group is a global research and consulting firm that focuses solely on data and analytics. Our experts specialize in data governance, self-service analytics, data architecture, data science, data management, and business intelligence.

Our clients say we are hard-working, insightful, and humble. It all stems from our love of data and our desire to help organizations turn insights into action. We are a family of continuous learners, interpreting the world of data and analytics for you.

Get more value from your data. Put an expert on your side. **Learn what Eckerson Group can do for you!**

Eckerson Group

RESEARCH CONSULTING ADVISORY

GET • MORE • VALUE • FROM • YOUR • DATA

About the Sponsor

Unravel Data radically transforms the way businesses understand and optimize the performance and cost of their modern data applications—and the complex data pipelines that power those applications. Providing a unified view across the entire data stack, Unravel's market-leading Data Observability platform leverages AI, machine learning, and advanced analytics to provide modern data teams with the actionable recommendations they need to turn data into insights. Some of the world's most recognized brands like Adobe, 84.51˚ (a Kroger company), and Deutsche Bank rely on Unravel Data to unlock data-driven insights and deliver new innovations to market.

To learn more, visit **www.unraveldata.com**.