

unravel™

APACHE  
Spark™

DATASHEET

# UNRAVEL FOR SPARK

AI-Powered Data  
Observability + FinOps  
Platform

Transform your Spark environment with intelligent automation, cost optimization, and performance improvement

## Overview

Unravel for Apache Spark is the industry's first AI-powered data observability and FinOps platform purpose-built for Spark workloads across any environment. By leveraging advanced machine learning, AI and predictive analytics, Unravel provides comprehensive visibility, automated optimization, and intelligent governance to help organizations maximize their Spark ROI—whether running on-premises, AWS EMR, Azure HDInsight, Google Dataproc, Kubernetes, or Databricks.

# 35%

**MORE WORKLOADS**

- Automatically optimize code, compute, and data by eliminating inefficiencies
- Achieve  $\pm 10\%$  budget forecast accuracy with predictive spend modeling

# 50%

**ENHANCED APP PERFORMANCE**

- Accelerate time-to-market for new AI initiatives
- Prevent pipeline bottlenecks before they impact production
- Streamline development with embedded CI/CD performance insights

# 90%

**LESS FIREFIGHTING TIME**

- Automated anomaly detection across data pipelines
- AI-enabled root cause analysis for rapid issue resolution
- Proactive alerts and auto-remediation capabilities

# 99%

**SLA ACHIEVEMENT**

- Predictive performance monitoring
- Automated cluster optimization and capacity planning
- Real-time resource contention detection

## Use Cases

### FinOps & Cost Optimization

- Real-time cost monitoring and optimization across all Spark resources
- Automated budget management and overage prevention
- Cross-team cost allocation and chargeback reporting
- Strategic pricing plan selection and reservation management

### Performance Engineering

- SQL query optimization for faster results and reduced resource consumption
- Workload performance tuning and capacity planning
- Bottleneck identification and resolution automation
- Service level agreement (SLA) assurance & monitoring

### Data Operations

- End-to-end data pipeline observability and quality assurance
- Proactive issue detection and automated remediation
- AI-driven root cause detection and analysis
- Operational efficiency improvement through automation

### Cloud Migration

- Legacy data warehouse migration planning and execution
- Post-migration optimization and performance validation
- Hybrid environment management during transition periods
- ROI measurement and business case validation

## Customer Successes

"Unravel cut our cloud data costs by 70% in six months—and kept them down."

- Financial Services Company

"Unravel helped us improve the platform resiliency and availability multiple fold."

- Retail company

"Equifax receives over 12 million online inquiries per day. Unravel has accelerated data product innovation and delivery."

- Equifax

"Unravel helped us reduce the time to identify and optimize top inefficient queries, cutting compute by 24%"

- Fortune Healthcare company

### Fortune Global 500 Bank

Implemented Unravel for developer pipeline evaluation, ensuring code performance before production release.

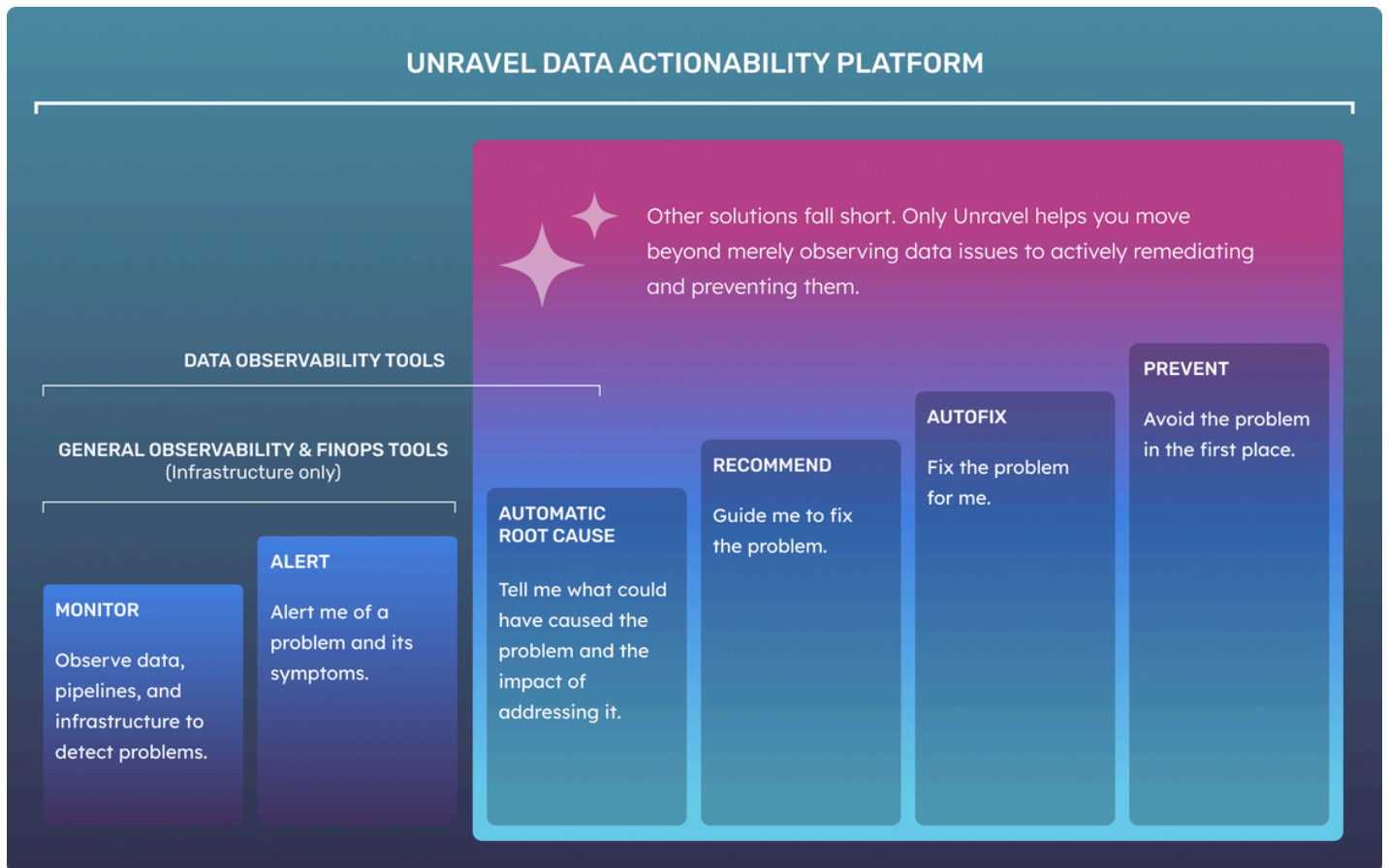
### Major Health Insurance Provider

Achieved optimal performance, reliability, and cost for business-critical data applications in development environments.

### Global Logistics Company

Used Unravel to upskill data teams at scale through CI/CD integration, ensuring quality and efficiency standards.





## Key Differentiators

### Beyond Traditional Observability

While traditional tools show "what" happened, Unravel's AI explains "why" it happened, "how" to fix it, and the automation to drive real results and move from reactive monitoring to proactive optimization.

### Purpose-Built for Spark

Unlike generic monitoring solutions, Unravel is specifically designed for the Spark architecture, understanding the nuances of jobs, stages, tasks, executors, driver coordination, shuffle operations, and memory management across all deployment platforms.

### Dynamic Analysis vs. Static

Unravel's recommendations are based on actual runtime performance and data characteristics, not just static code analysis.

### Financial + Technical Optimization

Unique combination of FinOps and performance optimization, ensuring both cost efficiency and technical excellence.

## Deployment Options

AWS Marketplace

Azure Marketplace

SaaS Deployment

### Security & Compliance

- SOC 2 Type II certified
- Enterprise-grade security controls
- Integration with existing security frameworks

## Core Capabilities

### Comprehensive Cost Management

- **Real-time Cost Visibility:** Granular breakdown compute costs, and resource consumption by user, team, job, and application
- **Predictive Forecasting:** AI-powered spend predictions with high accuracy for budget planning
- **Cost Optimization:** Automated recommendations for cluster right-sizing, code optimization, and data layouts
- **Chargeback & Showback:** Business-context cost allocation by project, department, or business unit

### Performance Optimization

- **AI-Driven Tuning:** Intelligent recommendations for Spark configuration, cluster settings, and resource allocation
- **Query Optimization:** SQL and code-level performance improvements with specific, actionable insights
- **Bottleneck Detection:** Proactive identification and resolution of performance constraints
- **Capacity Planning:** Automated forecasting of resource needs based on historical demand patterns

## Advanced AI Features

### Purpose-Built AI Insights Engine

- Trained on 50+ million data pipelines and queries
- Dynamic code analysis considering data layout, table structure, and metadata
- Runtime performance analysis for enhanced recommendations
- Continuous learning from your environment's patterns

### Intelligent Automation

- **Auto-Actions:** Automated remediation for common performance and cost issues
- **Smart Alerts:** Context-aware notifications with recommended actions
- **Guardrails:** Proactive prevention of cost overruns and performance degradation

### Data Quality & Reliability

- **End-to-End Monitoring:** Comprehensive observability across data pipelines and applications
- **Anomaly Detection:** ML-powered identification of data quality issues
- **SLA Management:** Automated tracking and alerting for data delivery commitments
- **Lineage Tracking:** Complete visibility into data flow and dependencies

### CI/CD Integration

- **Code Review Automation:** AI-powered analysis of code changes for performance and cost impact
- **Pre-Production Optimization:** Early detection and fixes of inefficiencies during development
- **DevOps Integration:** Seamless integration with GitHub, Azure DevOps, and other CI/CD tools
- **Developer Insights:** Real-time feedback on code efficiency and financial impact

# Spark Platform Support

## Universal Spark Coverage

- **On-Premises Deployments:** Full support for Spark on Hadoop (YARN), standalone clusters, and private cloud environments
- **AWS EMR:** Native integration with Elastic MapReduce for optimized Spark workload management
- **Azure HDInsight:** Comprehensive monitoring and optimization for Azure-based Spark clusters
- **Google Dataproc:** Complete observability across Dataproc Spark environments
- **Databricks:** Deep integration with Databricks on AWS, Azure, and GCP
- **Kubernetes:** Any flavor
- **Cloudera Data Platform:** Support for Spark workloads within Cloudera environments

## Technical Specifications

### Supported Platforms

- Apache Spark on AWS EMR, Azure HDInsight, Google Dataproc, Kubernetes, and Databricks
- On-premises Spark deployments (Hadoop YARN, standalone, Mesos)
- All major Spark versions (2.x, 3.x, 4.x)
- Hive metastore and other metadata repositories

### Integration APIs

- REST API for custom integrations
- Webhook support for external systems
- SAML/SSO authentication

### Monitoring Capabilities

- Real-time metrics collection
- Historical trend analysis
- Custom KPI tracking
- Automated reporting

## Comprehensive Coverage

- SQL, Python, Scala, and Java workload optimization
- Streaming and batch processing insights
- Machine learning pipeline monitoring
- Interactive notebook performance analysis

## STOP WASTING SPARK SPEND. ACT NOW WITH A FREE HEALTH CHECK

Optimize your Spark spend, boost performance, and increase ROI

Request your free report today at [unraveldata.com/health-check-demo/](https://unraveldata.com/health-check-demo/).

### About Unravel Data

Unravel Data is a data observability and FinOps platform that provides full-stack visibility, AI-powered recommendations, and actionable automations to help optimize the performance, cost, and reliability of modern data applications.

For more information, visit [unraveldata.com](https://unraveldata.com).