

unravel™

SOLUTION GUIDE

UNRAVEL DATA OPERATIONS GUIDE FOR AMAZON AWS

Planning, Migrating, and Managing
Modern Data Applications on AWS



TABLE OF CONTENTS

OVERVIEW

BIG DATA WORKLOADS ON AMAZON AWS	03
HOW UNRAVEL HELPS YOU GET THE MOST OUT OF YOUR DATA APPLICATIONS ON AMAZON AWS	04

01 PLAN

PLANING FOR AWS	05
DETERMINE YOUR OPTIMAL DEPLOYMENT MODE	05
PERFORMANCE BASELINING AND WORKLOAD ANALYTICS	06

02 MIGRATE

EXECUTING YOUR MIGRATION TO AWS	08
USE UNRAVEL AWS INSTANCE MAPPING TO OPTIMIZE PERFORMANCE – AND ROI	10

03 MANAGE

OPTIMIZING FOR COST	11
BOOST IT PRODUCTIVITY WITH RAPID, AI-DRIVEN TROUBLESHOOTING	11

HOW UNRAVEL WORKS

HOW DOES UNRAVEL WORK?	14
UNRAVEL AI-POWERED DATA OPERATIONS PLATFORM	15
UNCOVER UNDERSTAND UNRAVEL	16

UNRAVEL FOR AWS

STREAMLINED INSTALLATION FROM AMAZON AWS MARKETPLACE	17
UNRAVEL INTEGRATION WITH AMAZON EMR, ATHENA, AND REDSHIFT	18
UNRAVEL FOR AWS ARCHITECTURE	19

DATA OPPS IN THE CLOUD

UNRAVEL FOR DATA OPERATIONS IN THE CLOUD	20
UNRAVEL SUPPORTS THE TOOLS, SYSTEMS, AND ENVIRONMENTS YOU RELY ON MOST	21



BIG DATA WORKLOADS ON AMAZON AWS

There are compelling financial and operational reasons to run modern data applications on AWS:

Accelerated adoption of big data across the enterprise. Reduced spending and increased cost efficiency. Greater scalability, flexibility, and optimization. But, simply moving to AWS won't solve all of your problems, some challenges may stay with you and other, completely new challenges may arise for your cloud deployment.

Data teams need an AI-powered solution.

Today's data-driven applications rely on a multitude of technologies, which can make it difficult and expensive to get a clear picture of how these apps are performing. Developers, IT Operations, DevOps, and DataOps teams are under increasing pressure to operationalize their data and analytics pipelines. But to do so they need a unified approach

to understanding, planning, and optimizing their apps and infrastructure, wherever they're running. Unravel is the only unified, full-stack, AI-powered solution for tuning and troubleshooting your modern data applications in AWS.

HOW UNRAVEL HELPS YOU GET THE MOST OUT OF YOUR DATA APPLICATIONS ON AWS

Using advanced analytics, AI, and machine learning, Unravel helps your team proactively monitor and optimize modern data applications in both on-premises and cloud environments, like Amazon AWS, and hybrid and multi-cloud configurations.

Challenges Deploying Modern Data Applications in AWS

DIFFICULT TO PREDICT & PLAN
Which applications and datasets are best suited for AWS.

DIFFICULT TO MIGRATE
From on-premises hardware environments to virtual resources.

DIFFICULT TO CONTROL COSTS
And to get the best return on your cloud investment.

01 PLAN

Unravel helps you plan and organize your cloud migration by:

- Identifying which applications are best suited for a move to AWS
- Instantly map on-premises cluster usage to the optimal AWS instance types
- Data-driven capacity forecasting for intelligent budgeting and planning

02 MIGRATE

Unravel helps you transition your apps to Amazon AWS – and validate their performance – by:

- Baselining both performance and business metrics, to establish clear benchmarks
- Providing real-time visibility into what's working and what's not
- Validating post-migration success against performance baselines

03 MANAGE

Unravel helps you monitor and optimize data operations on AWS by:

- Determining the best use of permanent, transient, autoscaling, and spot instances
- Quickly detecting and resolving issues during each phase of migration
- Managing app performance across multiple platforms and environments, including cloud, on-premises, and hybrid cloud

PLANNING FOR CLOUD AWS

While moving traditional enterprise apps to AWS can be a fairly straightforward process, migrating big data applications to AWS is an order of magnitude more complex – and more challenging.

Unravel is designed to remove much of the complexity and difficulty of your migration process, using data-driven intelligence to identify ideal candidates for migration. Unravel can also help you make the right decisions about your deployment model and cloud resource management. And, by providing the insight you need to correctly size and provision AWS resources, Unravel ensures you get the best performance at the lowest cost.



ON-PREMISE



CLOUD & MULTI



HYBRID

Determine your optimal deployment model.

Unravel provides performance and resource profiles for your applications, so you can choose the best deployment model for your needs.

TYPE	STATUS	USER	APP NAME / ID	START TIME	DURATION	READ	WRITE	CLUSTER ID	QUEUE	GO TO
SPARK	SUCCESS	user3	com.unraveldata.apps.W... ofcCount	08/24/18 17:42:26	39s	2.8GB	0B	default	root.users.user...	WI
SPARK	KILLED	root	RealTimeSentimentMonit... application_1534794873	08/22/18 09:12:13	1h 46m 59s	1023.4MB	0B	default	root.users.root	
SPARK	SUCCESS	user3	RealTimeSentimentMonit... application_1534187915	08/16/18 10:07:04	1h 12m 26s	391MB	0B	default	root.users.user...	
SPARK	FAILED	root	RealTimeSentimentMonit... application_1534794873	08/22/18 03:35:08	42m 28s	728.8MB	0B	default	root.users.root	
SPARK	SUCCESS	root	RealTimeSentimentMonit... application_1534794873	08/22/18 00:10:16	42m 25s	941.4MB	0B	default	root.users.root	
SPARK	SUCCESS	root	RealTimeSentimentMonit... application_1534794873	08/21/18						

Identifying Ideal Migration Targets.

Unravel identifies applications that display bursts of processing activity, so you can take advantage of the aggregated compute resources that Amazon AWS provides.

Unravel has 3 recommendations to improve app efficiency.

TYPE	STATUS	ID	START TIME	DURATION	DATA I/O	# OF YARN APPS
MR	SUCCESS	1637	08/21/18 16:13:00	42.4 GB	5	7
MR	SUCCESS	1647	08/21/18 16:23:33	29.9 GB	5	7
MR	SUCCESS	1650	08/21/18 16:31:05	764.9 MB	2	7
MR	SUCCESS	1651	08/21/18 16:32:47	393.3 MB	2	7
MR	SUCCESS	1652	08/21/18 16:33:04	55.7 MB	2	7
MR	SUCCESS	1653	08/21/18 16:33:38	84.6 MB	3	7
MR	SUCCESS	1654	08/21/18 16:33:36	27.5 MB	3	7

```

select l_item_id
       ,l_item_desc
       ,l_data
       ,count(*) as store_sales_quantitycount
       ,sum(*) as store_sales_quantitysum
       ,count(*) as store_sales_quantitycount
       ,sum(*) as store_sales_quantitysum
       ,count(*) as store_returns_quantitycount
       ,sum(*) as store_returns_quantitysum
       ,count(*) as store_returns_quantitycount
       ,sum(*) as store_returns_quantitysum
       ,count(*) as catalog_sales_quantitycount
       ,sum(*) as catalog_sales_quantitysum
       ,count(*) as catalog_sales_quantitycount
       ,sum(*) as catalog_sales_quantitysum

```

Migrate failed or bottlenecked apps.

Unravel helps you locate applications portfolios that suffer failures and bottlenecks because on-premises clusters are running over capacity. Unravel can then help organize your cloud bursting efforts.

Unravel has 3 insights for this application.

SEVERITY	TIME	KEYWORD	MESSAGE
WARNING	08/16/18 03:08:41	yarn.YarnAllocator	Container killed by: OOM for exceeding memory limits. 1.4 GB of 1.4 GB physical memory used. Consider bootstrapping yarn.yarn.scheduler.memory.OomKilled.
WARNING	08/16/18 03:46:37	yarn.YarnAllocator	Container marked as failed: container_1534187915171_1348_01_000002 on host: lsgp233.unraveldata.com. Exit status: 56. Diagnostics: Exception from container-launch.
WARNING	08/16/18 03:46:40	yarn.YarnAllocator	Container marked as failed: container_1534187915171_1348_01_000018 on host: lsgp233.unraveldata.com. Exit status: 56. Diagnostics: Exception from container-launch.

Resource usage graph showing processCpuUsed over time (08:10:15 to 10:00:00).

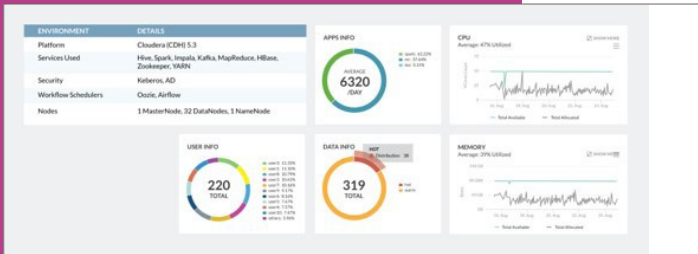
Identify apps with variable resource usage.

Unravel identifies cloud migration candidates from applications that have a wide variation of cluster resource usage – either from seasonality or variability in the size and number of datasets and users.

Performance Baseline and Workload Analytics.

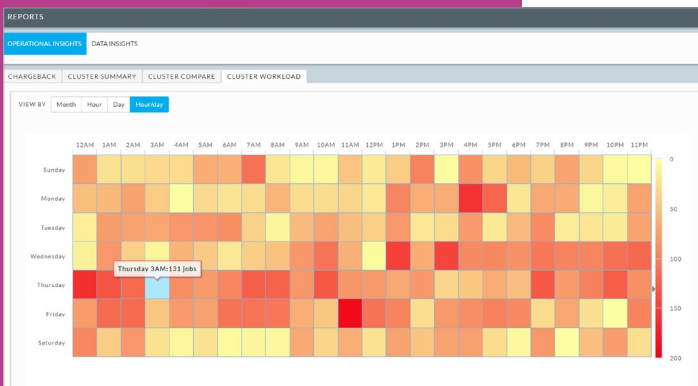
Migrating to AWS is challenging enough, but moving distributed data applications to AWS adds even more difficulties. To effectively migrate Spark data pipelines from physical to virtual data centers, you need deep data and intelligence.

For example, your choice of Amazon AWS server instances is critical to the success of your migration. Unravel can infer the seasonality of your big data workloads and deliver recommendations for optimal server instance sizes – in minutes instead of hours or days.



Create On-Premises Cluster Discovery Reports in minutes.

Unravel can provide detailed reports on your on-premises clusters, including total memory, disk, and the number of hosts and cores used. This Cluster Discovery Report also delivers insights on cluster topology, running services, OS version, and more. And resource usage heatmaps can be used to determine your unique needs for AWS.



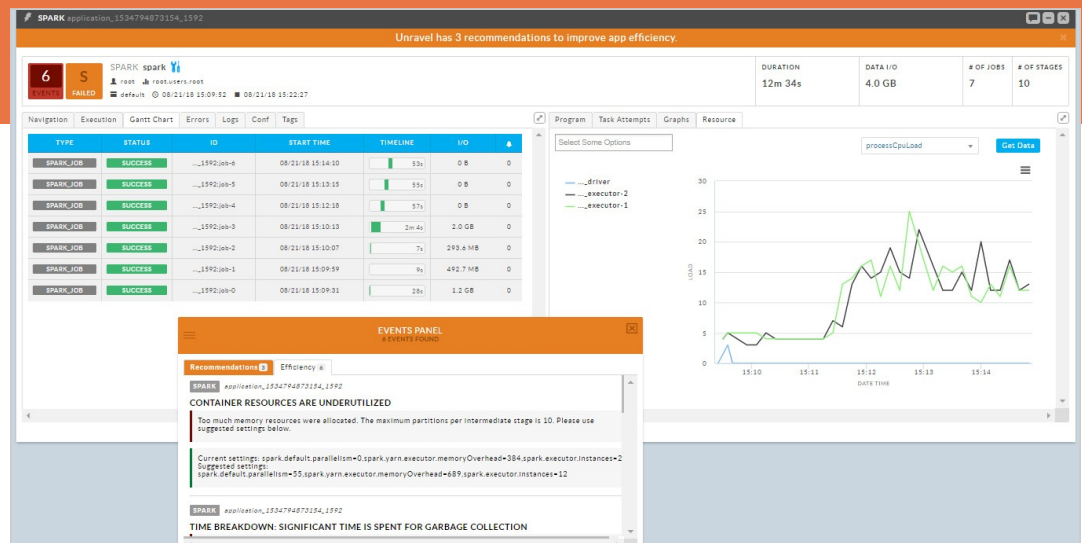
Gain key app usage insights from cluster workload analytics.

Unravel can highlight application workload seasonality by user, department, application type, and more – to help you best make use of AWS resources. This type of reporting can also aid in taking advantage of permanent, transient, autoscaling, and spot instances to maximize your ROI on AWS expenses.

EXECUTING YOUR MIGRATION TO AWS

Unravel makes it easy to baseline performance, for a before-and-after comparison of how your applications are performing on AWS – and tangible recommendations for improving that performance.

For instance, a baseline comparison reveals that this app is 8x slower on AWS. Using AI-powered intelligence, Unravel can uncover automated fixes and actionable recommendations to help the application start meeting its SLA again. These recommendations can include cluster configuration changes, host and cloud instance sizing, and application parameter changes for all parts of the data stack.

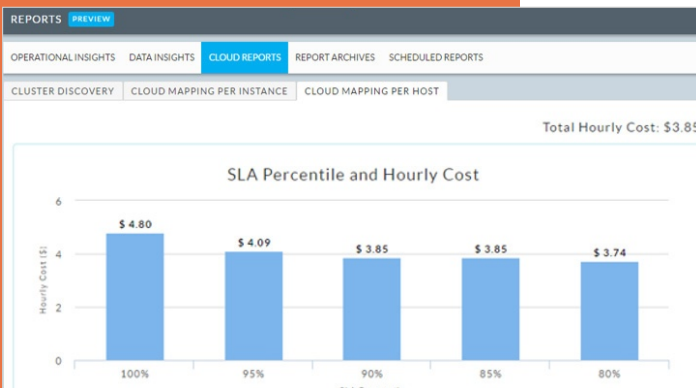


	Successful	Failed/Killed	SLA %	Cost
Pre Migration	550	0	100%	\$4910
Post Migration	510	40	89%	\$6820

APP PERFORMANCE - PRE				APP PERFORMANCE - POST			
App Name	Status	Duration	Cost	App Name	Status	Duration	Cost
Spark 1	S	3m 4s	\$2.50	Spark 1	S	2m 58s	\$2.45
Spark 2	S	14m	\$12.50	Spark 2	S	17m	\$14.35
WF 1	S	4m 14s	\$22.10	WF1	F	-	\$14.10

Quickly validate migration with baseline comparisons.

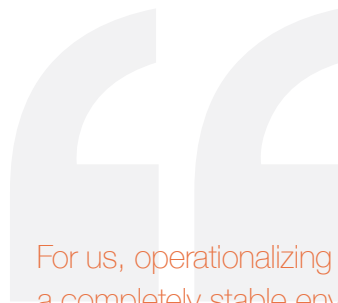
Typically, manually collecting performance data from your current environment can take days or weeks. With Unravel, it takes minutes. Unravel can then quickly collect runtime metrics from your cloud environment to easily compare how performance has changed.



Unravel Cloud Instance Mapping.

Once you've started creating On-Premises Cluster Discovery Reports, Unravel can use that data to map your data's center physical server environment to the ideal virtual server instances in AWS. Unravel provides three different profiles for mapping clusters to AWS:

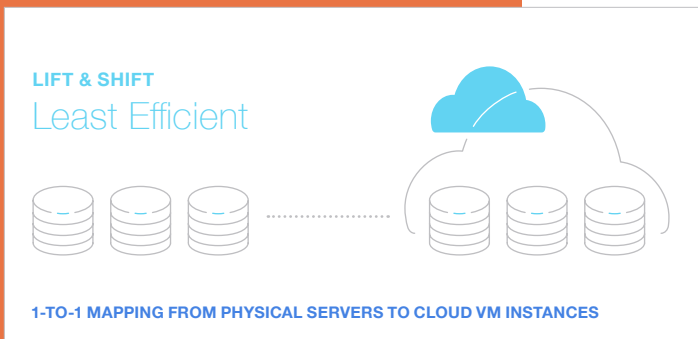
- Lift and Shift Profiles
- Cost Reduction Profile
- Workload Fit Profiles



For us, operationalizing big data means creating a completely stable environment that allows us to plan, report, and optimize app performance to support our business needs. Unravel knows where those weaknesses are – and how to fix them.

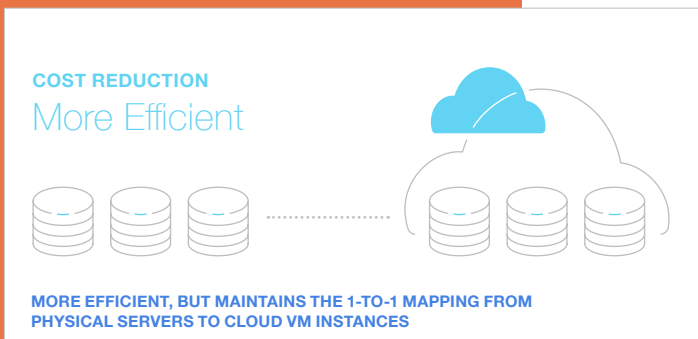
— **HEAD OF OPERATIONS,**
LEADING NORTH AMERICAN TELECOMMUNICATIONS PROVIDER

Use Unravel AWS Instance Mapping to optimize performance – and ROI.



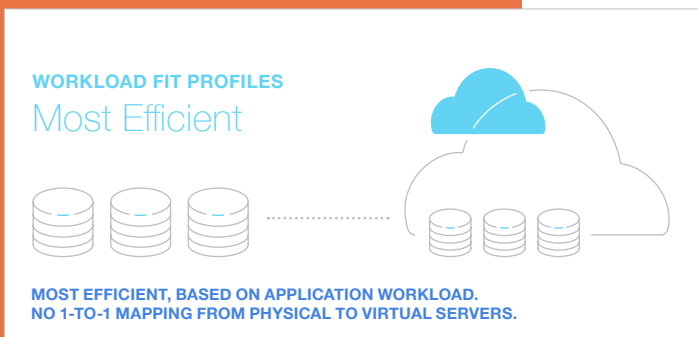
Lift and Shift Profiles

provide one-to-one mapping of physical hosts to cloud instances, matching those instances to the original server configuration as closely as possible. This approach isn't always optimal because it doesn't take your app workloads into consideration, and therefore doesn't account for the cost savings that moving to AWS can provide.



Cost Reduction Profiles

also look at your typical application workloads to provide the most cost-effective instance recommendations for minimizing wasted capacity and overprovisioning. Cost-reduction reporting is always less expensive than lift-and-shift, but it still utilizes one-to-one mapping, which can mean some remaining overprovisioning of cloud resources.



Workload Fit Profiles

take into account data collected over time from the on-premises environment, making recommendations for instance types based on the actual workload of applications running in your data center. These recommendations will be based on the VCore, memory, and storage requirements of your typical runtime environment. Because of this additional intelligence, workload-fit reporting can often be the most cost-effective.

OPTIMIZING FOR COST

The decision to migrate your modern data applications to AWS is the result of significant financial calculations made by your CIO and CFO.

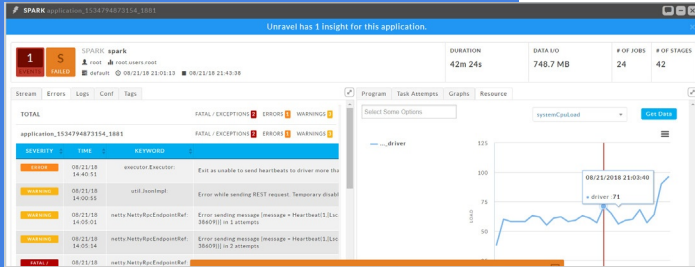
But, getting the most out of your cloud budget requires calculations of another sort: Using AI, machine learning, rules engines, and advanced analytics to optimize how you use your team and resources.

The screenshot displays three alerts from the Databricks console for a Spark application (application_1482293458178_0368):

- LARGE IDLE TIME FOR EXECUTORS:** Executors had no tasks to run for 42.0 minutes.
- OPPORTUNITY FOR RDD CACHING:** 9.0 minutes spent recomputing RDDs in the application. A suggestion is provided: "Adding a cache() statement before count at PetFoodAnalysisCaching.scala:129 can save up to 9.0 minutes. Caching with StorageLevel.MEMORY_AND_DISK_SER is recommended."
- CONTAINER RESOURCES ARE UNDERUTILIZED:** Too much memory resources and cpu resources were allocated. Please use suggested settings below. Current settings: spark.executor.instances=2,spark.executor.memory=3221225472. Suggested settings: spark.executor.instances=5,spark.executor.memory=2381163942.

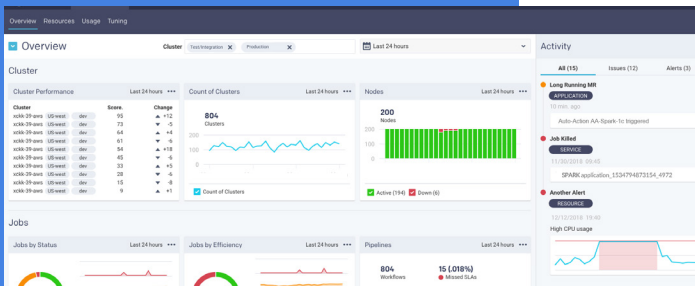
Boost IT productivity with rapid, AI-driven troubleshooting.

Unravel uses AI and machine learning to cut through the complexity of modern data pipelines and assist developer and operations teams in finding the root cause of issues at both the application and platform level.



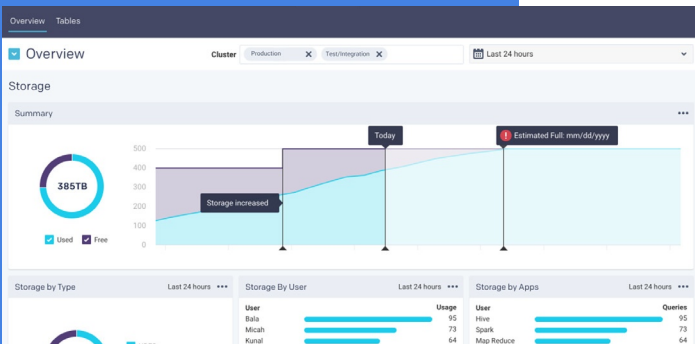
Increase capacity with auto-tuning.

Unravel can significantly reduce resource consumption – making room for more jobs – by auto-tuning applications, as well as automatically finding and eliminating resource wastage.



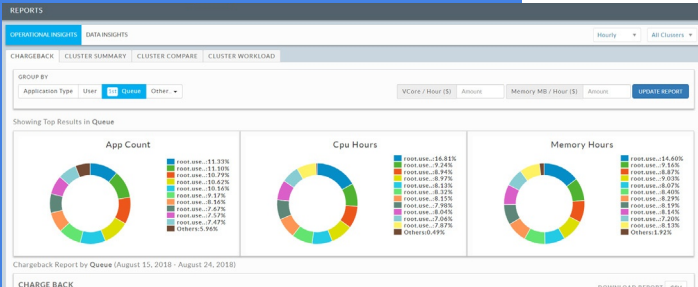
Rightsize the allocation of cluster resources.

Unravel automatically sizes containers and tunes cluster configurations for optimal throughput in.



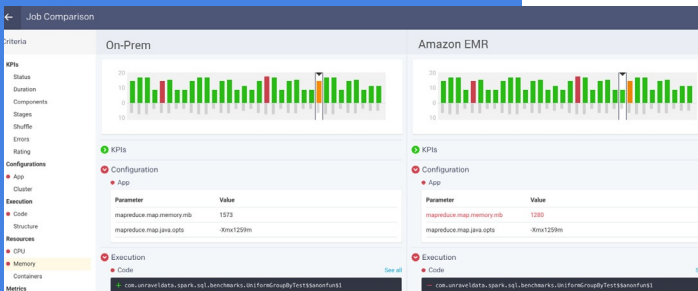
Cut storage costs with multi-temperature data tiering.

Unravel can identify hot, warm, and cold data, helping you realize additional costs savings by moving cold data to cheaper storage services like Amazon Glacier.



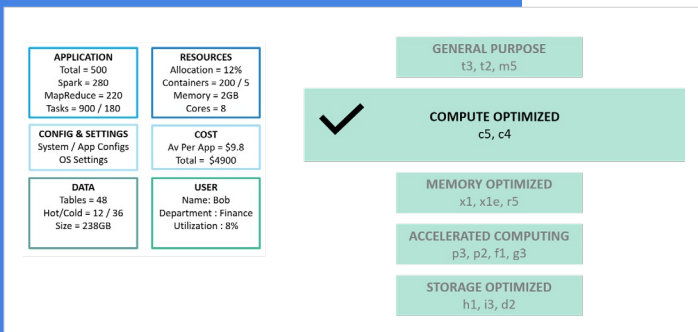
Facilitate full accounting with chargeback reporting.

Unravel allows you to generate chargeback reports for multi-tenant cluster usage costs, organized by application type, user, queue, and user-defined tags like department and location. These reports can be used to provide a complete accounting of your AWS migration.



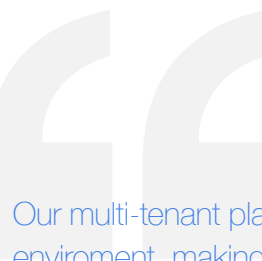
Uncover potential cost savings.

Unravel provides detailed reports on cloud operating costs, revealing potential cost savings for AWS deployments.



Correlate different datasets to determine rightsize capacity.

Collecting and correlating workload metadata, Unravel can match your application workload profile to the ideal AWS instance type, providing the most cost-effective options for getting the job done.



Our multi-tenant platform is a very complex environment, making tooling an ongoing challenge for both our dev and ops teams. Unravel gives us the 360° view we need to get proactive with troubleshooting and supporting our SLAs."

HOW DOES UNRAVEL WORK?



Unravel is a unified, AI-powered, full-stack solution for understanding and improving the performance of your data pipelines.

Designed to quickly and automatically optimize large-scale data applications, Unravel actively collects data from across apps and pipelines, then uses AI, machine learning, and predictive analytics to analyze the current state of your ecosystem, making recommendations and taking action to improve application performance.

At the highest level, Unravel allows you to:

uncover

Capture real-time metadata from across all of your applications, systems, and infrastructure, for 360° full-stack visibility.

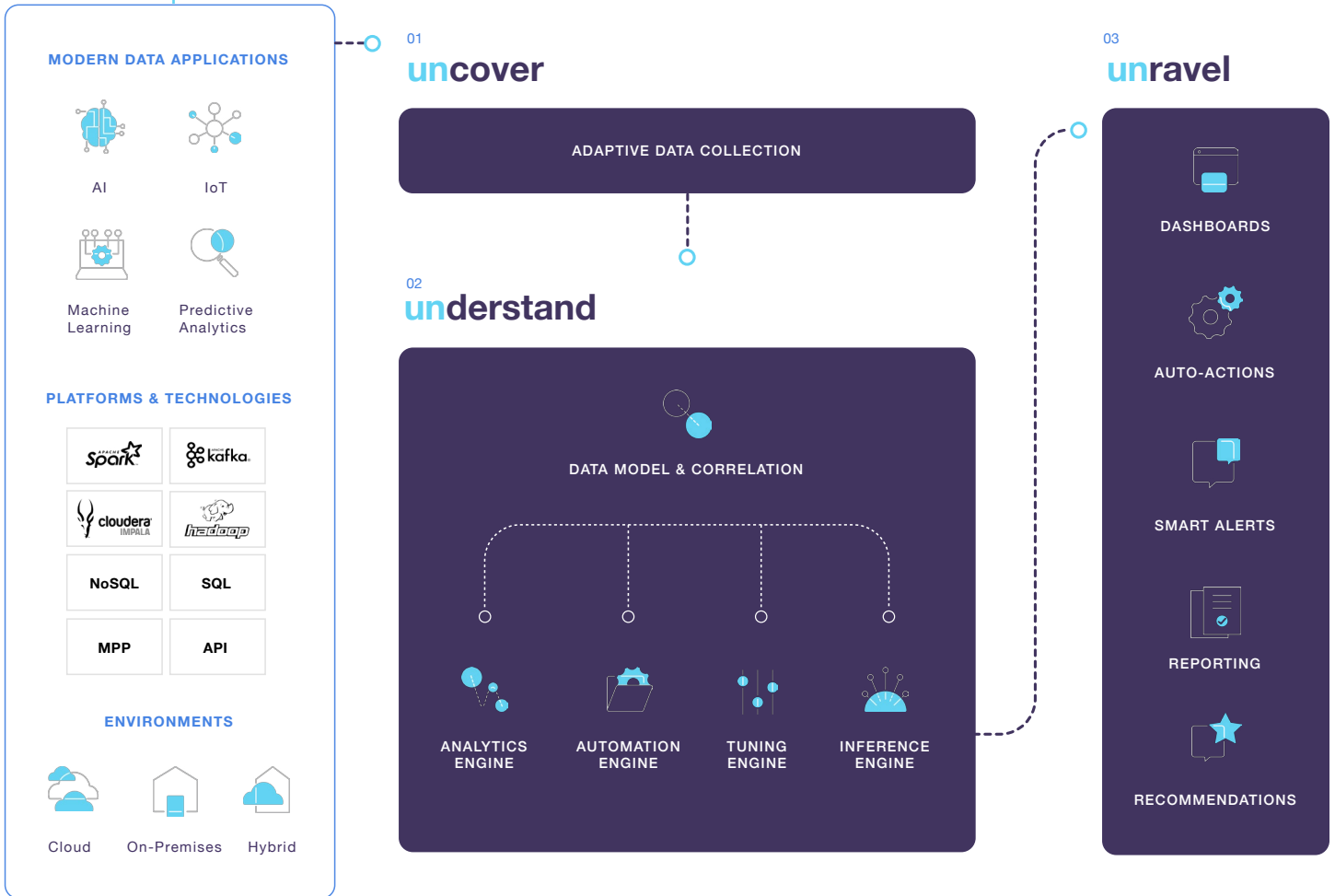
understand

Analyze and correlate this metadata to model your pipelines from end-to-end, creating a single, unified view for monitoring performance.

unravel

Get AI-driven intelligence and actionable recommendations for optimizing your apps and ecosystem, including smart alerts, dashboards, and automated actions.

Unravel AI-Powered Data Operations Platform



01

uncover

Your big data stack is so massive and complex that it can be hard to see how all the components fit together, impact each other, and affect performance.

Lightweight and agentless, Unravel uses native APIs and micro-sensors to non-intrusively capture metadata from every element of your stack, for complete visibility.

The full range of data collected by Unravel includes metrics on:

- Applications
- Infrastructure (including components like Spark, Kafka, Hadoop, NoSQL, and beyond)
- Analytics, machine learning, IoT, and business intelligence workloads
- On-premises, in the cloud (including AWS and Azure), and hybrid and multi-cloud environments
- Cloud metadata including vCORE load, memory usage, IO, billing, and instance

02

understand

Unlike other monitoring tools, which merely aggregate metrics in charts for you to figure out, Unravel provides the full context you need to plan, manage, and improve performance.

Using AI, machine learning, and advanced analytics, Unravel quickly analyzes and correlates all of the metadata it collects to create a dynamic data model of performance across your stack, apps, resources, datasets, and users.

By providing this unified view – how each element is functioning, how different components depend on and impact each other – Unravel helps you monitor and understand performance like never before.

03

unravel

Unravel doesn't just monitor performance – it offers tangible, actionable recommendations for tuning, troubleshooting, and optimizing performance: Clear-cut code you can use, settings you can tweak, resources you can reallocate.

With policy-driven automation, Unravel can also use this intelligence to automatically take action, based on rules you specify. These Auto-Actions include:

- Auto-tuning slow apps and stack components
- Killing rogue processes and applications
- Moving jobs from one queue to another based on priority or service level violation
- Executing custom scripts, processes, and applications via HTTP callouts

Smart alerting delivered via email, Jira, Slack, and PagerDuty. With Unravel, every member of your team is a big data expert, empowered with AI to improve the performance of your apps and ecosystem to ensure optimized service levels.

UNRAVEL FOR AWS

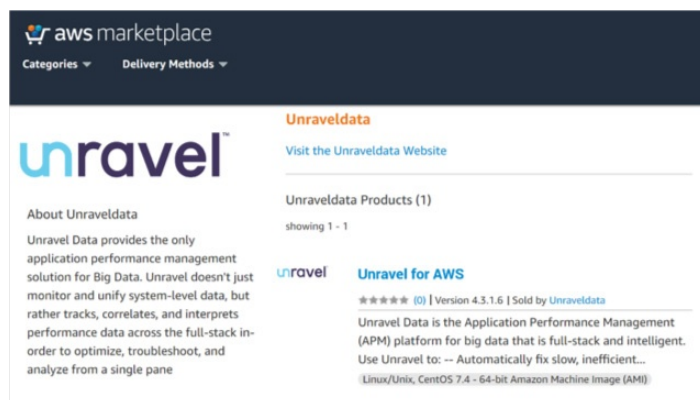
Unravel is a comprehensive platform for automated performance tuning, root cause analysis and troubleshooting, capacity forecasting and planning, and migration and validation of modern data applications on Amazon AWS.

The Unravel platform is essentially the same product in the cloud or on premises, but Unravel has partnered with Amazon to provide many custom integrations with services like the AWS marketplace, Amazon EMR, Amazon S3, Redshift, and Athena.

Streamlined Installation from Amazon AWS Marketplace

Unravel is available today on Amazon AWS marketplace and can run directly on AWS, or as part of an Amazon EMR cluster. Users have a free 30-day, full-featured trial period and can bring your own license (BYOL) to run Unravel.

Installation and configuration of Unravel takes less than an hour (although more complicated customer cluster environments may take a bit longer).



Unravel Integration with Amazon EMR, Athena, and Redshift

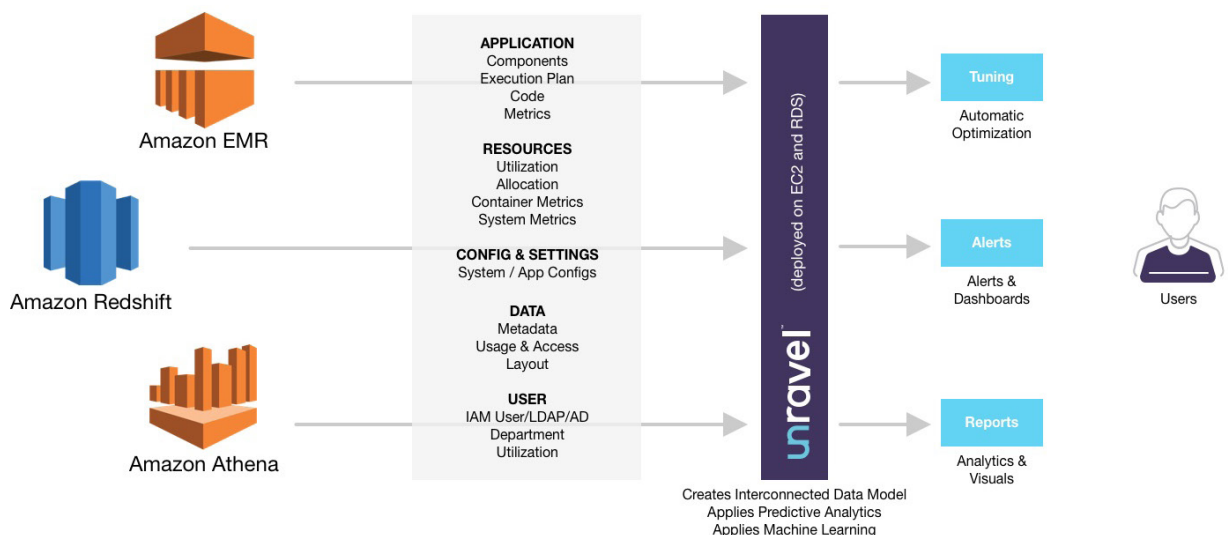
Unravel provides native integration to Amazon AWS services like EMR, Athena, and Redshift using Unravel's standard mechanisms of adaptive data collection, creating and maintaining the interconnected Unravel data model, and then uses predictive analytics, machine learning and AI tune and troubleshoot the Amazon AWS environment.

For Unravel AWS environments, Unravel collects the data needed to understand the root cause of failures and slowdowns. This includes:

- **Application metadata** around app components, execution plan, application source code and runtime metrics
- **Cluster resource metrics** like utilization levels, resource allocation per app/job, container metrics and system/cloud metrics
- **Configuration and settings** at both the application, cluster and system levels.

- **Data metrics** including runtime metadata, usage and access, and data layout/format
- **User metrics** including directory services like IAM, LDAP, and Active Directory, and additional user related metadata such as department/organization, user name, and per user resource utilization

Unravel in turn uses this interconnected data model to provide alerting, auto-tuning, automated troubleshooting, reporting, and auto-actions to kill jobs, move jobs to a different queue, or notify team members via slack, PagerDuty or email.

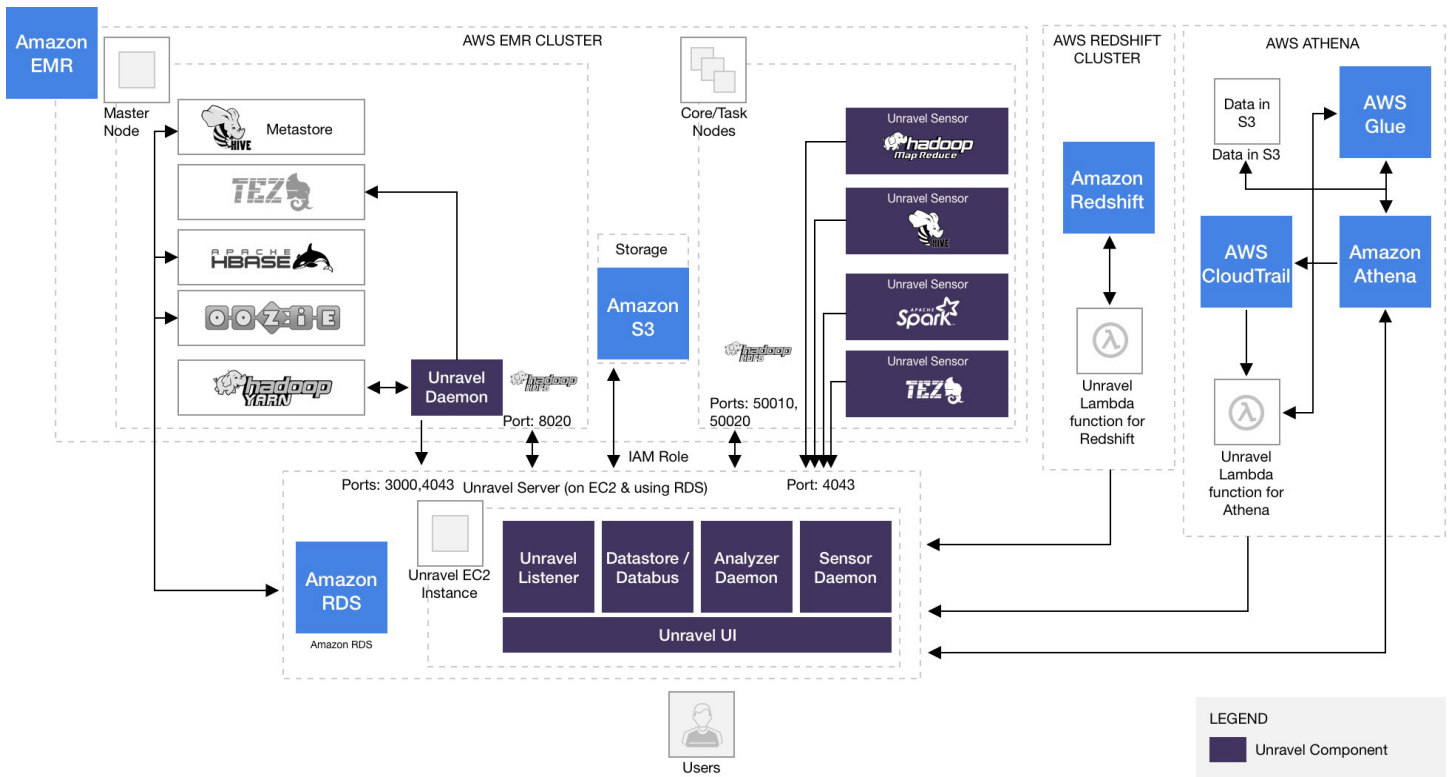


Unravel for AWS Architecture




The diagram displays the high level architecture of how Unravel integrated with the Amazon AWS environment, with Amazon EMR, Redshift and Athena prominently featured. The diagram also shows the big data stack components that are part of the Unravel solution including native integration of

Hadoop (HDFS, Hive, HBase, YARN, MapReduce, etc), as well as Spark, Kafka, Oozie and more.

For a more complete view into the Unravel product architecture, please refer to the [Unravel Technical Whitepaper](#).



UNRAVEL FOR DATA OPERATIONS IN THE CLOUD

Supported Cloud Environments

A full-featured trial version of Unravel is available as on both Amazon and Azure clouds.

Unravel for Amazon AWS and EMR

Unravel is available on Amazon AWS and Amazon EMR, supporting the following cloud services:

Unravel for Microsoft Azure

A Microsoft Co-Sell Partner, Unravel is available in the Azure Marketplace and supports the following cloud services:

Unravel supports the tools, systems, and environments you rely on most.

Big Data Ecosystem

SYSTEMS AND ENGINES



WORKFLOW SCHEDULERS



PLATFORMS



Environment

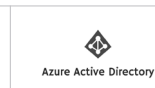
MICROSERVICES



INFRASTRUCTURE ENVIRONMENTS



SECURITY AND ACCESS CONTROL



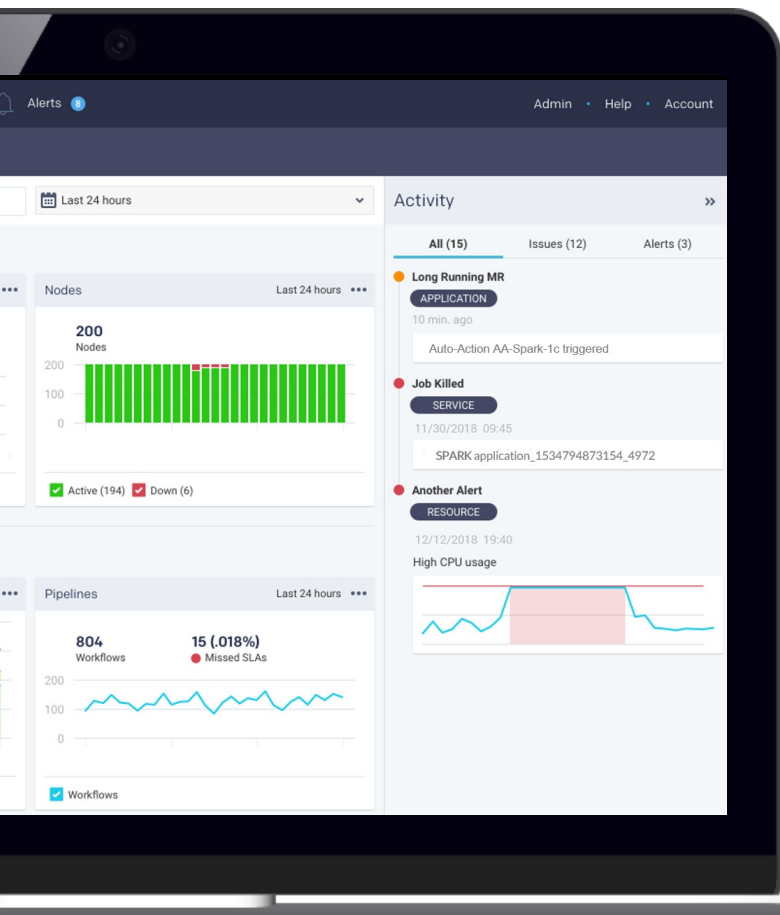
Other Tools

MONITORING



COLLABORATION





READY TO MIGRATE AND OPTIMIZE YOUR AMAZON AWS DATA WORKLOADS?

START YOUR FREE TRIAL →

About Unravel

Unravel radically simplifies the way businesses understand and optimize the performance of their modern data applications – and the complex pipelines that power those applications. Providing a unified view across the entire stack, Unravel's AI-powered data operations platform leverages AI, machine learning, and advanced analytics to offer actionable recommendations and automation for tuning, troubleshooting, and

improving performance – both today and tomorrow. By operationalizing how you do data, Unravel's solutions support modern big data leaders, including Kaiser Permanente, TIAA, Adobe, Deutsche Bank, Wayfair, and Neustar. The company is headquartered in Palo Alto, California, and is backed by Menlo Ventures, GGV Capital, M12, Data Elite Ventures, and Jyoti Bansal. To learn more, visit unraveldata.com.

unravel

unraveldata.com | hello@unraveldata.com

© Unravel. All rights reserved. Unravel and the Unravel logo are registered trademarks of Unravel. All other trademarks are the property of their respective owners.