unravel

Unravel Newsletter: August 2019

Greetings from Unravel,

Current trends from various sources like this and this as well as conversations with you all clearly indicates the trend of big data workloads moving to the cloud. More and more organizations like yours, already have or are looking to either completely migrate to the cloud or run significant portions of their big data workloads in the cloud.

Unravel provides full-stack visibility and AI-powered guidance to help customers understand and optimize the performance of their data-driven applications and monitor, manage and optimize their big data systems. This applies as much to workloads in the cloud as it does to big data clusters on-premise. Specifically, for the cloud, our goal is to cover the entire gamut below:

- IaaS: CDH/HDP/MapR deployed on cloud VMs where your modern data applications are running.
- PaaS: Managed Hadoop, Spark Platforms like Amazon Elastic Map Reduce (EMR), Azure HDInsight, Google Cloud Dataproc etc.
- Cloud-Native: Products like Amazon Redshift, Azure Databricks, AWS Databricks, Microsoft SQL Datawarehouse etc.
- Serverless: Ready to use, no setup needed services like Amazon Athena, Google BigQuery, Google Cloud DataFlow etc.

In addition, Unravel also has an offering to help in the phase before workloads are in the cloud – meaning the planning for Migration to the cloud.

In this month's newsletter, we focus on some recent updates about Unravel Cloud related offerings. We'll continue with some community highlights in the area of Data and Analytics and include links to recent recordings and upcoming events that you may find interesting.

Announcing the Cloud Migration Assessment Offering

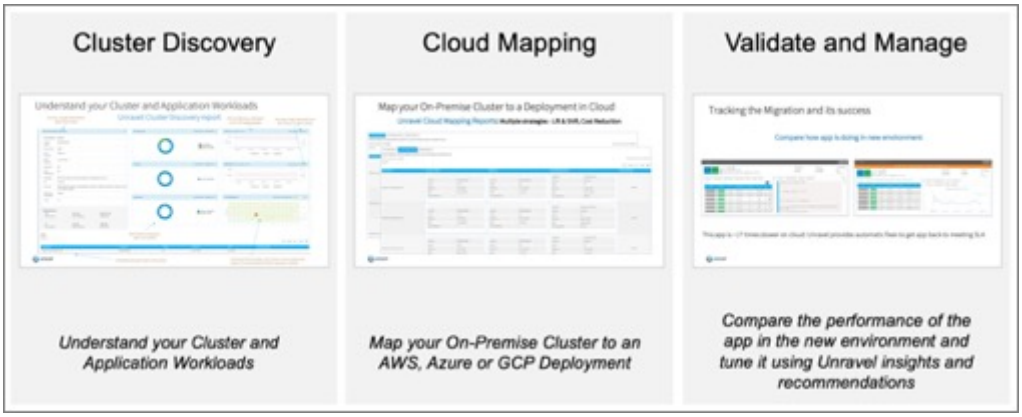
We recently announced the Unravel Cloud Migration Assessment

Offering to help organizations like yours plan the migration of your on-premises big data workloads to the cloud (Azure, AWS or GCP). Please refer to the announcement for more details.

In general, the planning and process of moving to the cloud is quite complex. It is a technically challenging and difficult task to derive the right pieces of information from the existing on-premises environment, combine those with the details of the cloud offering you aim to migrate to, correlate and synthesize all that information and come up with the optimal cloud topology that will meet the business goals set forth by an organization (e.g. minimizing risks or minimizing costs).

Unravel's offering takes all the guesswork and manual process out of this. Unravel begins with discovering your cluster, identifying workloads suitable for the cloud, then finds the optimal cloud topology based on your business strategy (minimize risk/minimize costs) and computes the anticipated hourly costs. This greatly reduced the time and complexity of planning the migration to the cloud, prevents errors and also reduces the costs that would be incurred for the migration process as well.

Check out a sample Unravel Migration Assessment report here.



Summary: Unravel Cloud Migration Workflow

Learn more about this feature in the Migration Planning Section of the Unravel User Guide.

Also refer to this blog Migrating big data workloads to Azure HDInsight by Arnab Ganguly, Senior Program Manager, Microsoft, where he discusses how Unravel brings a host of services towards providing unified visibility and operational intelligence to plan and optimize the migration process onto Azure HDInsight.

Unravel for Azure HDInsight and Amazon EMR

As some of you may be aware, Unravel provides all the same benefits and value propositions for Azure HDInsight and Amazon EMR as we do for on-premise CDH/HDP/MapR deployments.

- Run applications more reliably, improve performance (improve SLA)
- Run applications more reliably (lower costs)
- Root Cause Analysis - Detect, troubleshoot and fix issues faster (lower MTTR)
- Operational Intelligence – visualize and derive insights on usage, charge back/show back, understand resource usage, achieve overall cluster optimization etc.

All of this functionality is supported by a single Unravel Deployment across all your long-running and/or transient (ephemeral) clusters! So you get insights and recommendations from Unravel that span across clusters. As an example, you can easily see chargeback/show back reports for all your clusters or for a set of clusters or simply for a single cluster you select. Unravel recommendations can help you improve performance and reduce the duration for applications to run. If such an application is executed on ephemeral clusters, this directly results in cost reduction.

Check out this blog post by George Demarest, Senior Director of Product Marketing, Unravel and Abha Jain, Director Products, Unravel: Meeting SLAs for Data Pipelines on Amazon EMR With Unravel. This blog, first published on the Amazon Startup blog, discusses how a well-known global media analytics company is using Unravel to support their data operations (DataOps) on Amazon EMR to establish and protect their internal service level agreements (SLAs) and get the most out of their Spark applications and pipelines.

We continue to invest aggressively into Unravel offerings for these platforms given the usage and adoption trends we are noticing with our current customers as well as in the market in general. You will notice frequent releases for these platforms here as we roll them out to ensure support for the latest versions of the cloud platforms, as well as add new features and functionality.

Early Access Program for Unravel for Azure Databricks

We are excited to announce the Early Access Program for the Unravel offering for Azure Databricks. This provides **Application Performance Management & Operational Intelligence** for Azure Databricks.

In this release, we focused on Data Engineering use cases. We Analyze & Optimize workloads on Job Clusters across multiple Databricks Instances & Workspaces.
Features and functionality:

- Faster Diagnosis & Resolution of Application Issues
- Higher Cluster Resource Efficiency
- Unified View Across Workspaces & Instances
- Ensuring Reliable Apps in Production
- Usage Breakdown & Trending
- Speed up Applications
- Single Pane of Glass for Apps, Resources, Data, & Users

All the Databricks Job types: Notebook, Jar, spark-submit, and Python Task can be monitored by Unravel. This functionality is available on a single Unravel Deployment across all your Databricks Instances & Workspaces.

You are welcome to get in touch with us to join the Early Access Program and try this out! As an early access customer, your feedback will carry utmost weight in how this offering shapes up.

Stay tuned for more exciting announcements in future!

Conference Sessions & Webinar Recordings

- Check out the video from the talk by Kunal Agarwal, Unravel CEO at Venture Beat Transform 2019, San Francisco. Kunal discusses the complexity of modern data applications and how Unravel is simplifying data operations.
- Watch this webinar recording by George Demarest, Unravel Senior Director of Product Marketing - Understanding DataOps and its Impact on Application Quality with DevOps.com and Unravel. Learn about what DataOps is, and how critical good data ops is to the integrity of your application. Intelligent APM for your data is critical to the success of modern applications.

Upcoming Events & Webinars

- How to Optimize Spark Data Pipelines on Azure Databricks - Join Unravel expert Aengus Rooney to develop an understanding of the performance dynamics of modern data pipelines and applications. In this session, you will learn about uncovering and understanding the key datasets, metrics, and best practices

needed to develop mastery with Spark performance management on Azure Databricks.
- Join us the Unravel Booth at the Strata Data Conference New York, in New York, Sept 23-26.
  Join us at the Unravel Booth at AWS Summit Global Summit Toronto, Toronto, Sept 20.

Community Highlights

Some interesting articles in the area of Data and Analytics:

- Lumen: Custom, Self-Service Dashboarding For Netflix - Netflix generates a lot of data. One of the ways that they gain useful insights is by visualizing that data in dashboards which allow them to comprehend large amounts of information quickly. Here's how dashboards are used to track and chart key business metrics, compare the results of experiments, monitor real-time data and much more
- OctSQL is a tool for querying multiple data sources—whether remote DBs or local files. OctoSQL is written in golang, and configuration of data sources is via a simple yaml configuration format.
- LinkedIn has open sourced Brooklin, their tool for streaming data between systems. Brooklin is a multi-tenant, dedicated service with dynamic provisioning/management (via REST endpoints). At LinkedIn, it's used to stream data between streaming systems (e.g. Kafka and Azure EventHubs as well as Kafka to Kafka), change data capture, and more. It currently supports MySQL, Cosmos DB, and Azure SQL as data sources as well as Kinesis, Cosmos DB, and Couchbase as destinations.

Resources

- Learn more about Unravel.
- Online Product Demo
- Unravel Partners
- Unravel Product Releases and Documentation
- Unravel Datasheet
- More Unravel News