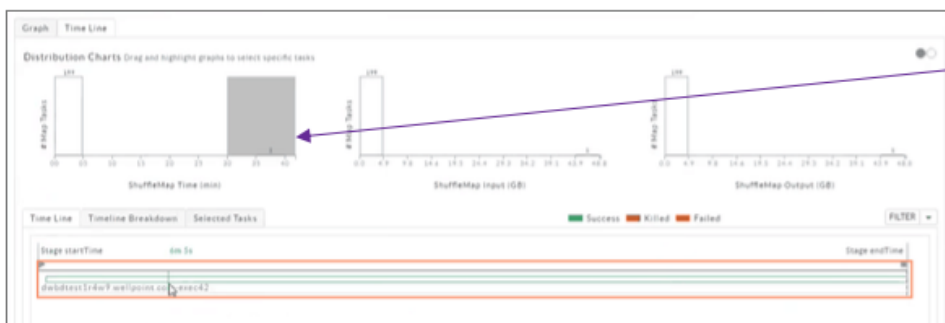# Unravel Newsletter: July 2019

Greetings from Unravel,

Hope you had a fun and refreshing July 4th weekend!

In this month's newsletter we discuss another scenario in which Unravel can help you optimize your Spark Applications. We then touch upon some recent Unravel product enhancements. 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.

Using Unravel to tune Spark Data Skew and Partitioning
In Spark, it's very important that the partitions of an RDD are aligned with the number of available tasks. Having too few partitions leads to less concurrency, processing skew, and improper resource utilization, whereas having too many leads to low throughput and high task scheduling overhead.

Using Unravel, you can identify the Stage in a Spark application that is the bottleneck in the execution by checking out the Gantt Chart View. You can then navigate to the specific stage and inspect each task within that stage. The timeline view can reveal possible skew issues.

All but one tasks completed in < 5 min. One takes 35-40 min to complete, indicating skew

Unravel indicating a skew issue in a Spark Application

Unravel can provide valuable recommendations for optimizations in several cases where join key(s) or group by key(s) are skewed.

Sam Lachterman, Manager, Solutions Engineering at Unravel, recently published a blog discussing best practices for how to use Unravel to tackle issues of data skew
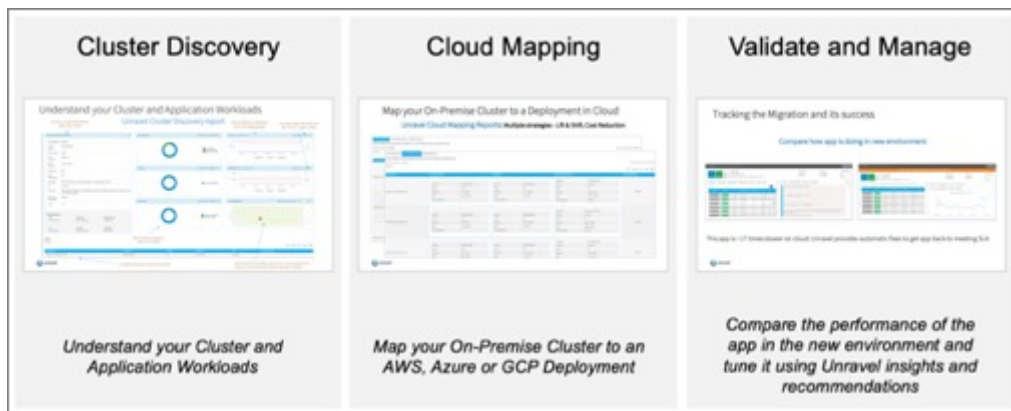
and partitioning for Spark applications: Using Unravel to tune Spark Data Skew and Partitioning.

Also check out the Spark Application Program Manager Section in the Unravel User Guide to help you get started with monitoring and optimizing Spark Applications running in your cluster.

Next, a glimpse into some recently released features and enhancements. Many of these came as feedback from Unravel customers like yourself. Please reach out to your Unravel Customer Success contact about joining the Unravel Product Advisory Council (UPAC) to get a front seat in helping us shape our roadmap.

Recent Product Enhancements

- **Migration Planning**
    - Unravel helps you plan your migration from an on-prem cluster to the cloud. Unravel begins with discovering your cluster, then finds the optimal cloud topology based on your business strategy (minimize risk/minimize costs) and computes the anticipated hourly costs.

Summary: Unravel Cloud Migration Workflow

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

- **Enhancements to Top X ReportMigration Planning**
    - This report lists the top X apps, by app type, for the categories - Longest Duration, Highest Dis I/O, Highest Cluster Usage, Highest CPU Usage and Highest Memory Usages. Here's a sample snippet of such a report:

A snippet of the Top-X Report

- We made the following enhancements to this report recently:
    - Added support for Impala apps
    - Custom tags now displayed for associated apps
    - These reports can now be filtered based on User/Queue/Tags

Top-X Reports - Enhanced Filtering

- **Impala**
  - Added support for running Impala queries.
  - Running Impala queries now have the AutoActions metric

- **Miscellaneous**
  - Enhanced support for Hive on Spark (Hive and Spark can now be linked)
  - RBAC - applications can be filtered by the real user by setting the property
  - Added support for YARN log aggregation IndexedFormat

In addition, we delivered several bug fixes and improvements aimed to make Unravel even more valuable for our customers. Check out release notes here for a complete list of enhancements and fixes for all releases.

Stay tuned for some big product announcements coming in the next months!

Announcing the Unravel Partner Program

We recently announced the Unravel Partner Program with a powerful list of brands including Informatica, Attunix, Dell, Clairvoyant LLC, and RCG Global Services. We're excited to fuel the demand for cloud migration and data operations initiatives. Check out the release for full details: http://bit.ly/2YcbUtr.

Recordings

- Watch a demo from Dave Berry, Senior Solution Engineer at Unravel Data, as he takes you through an AI/ML based approach to APM applied to data applications on any infrastructure - whether it be cloud, on-premise, or a combination of the two: Moving Big Data Pipelines to the Cloud: Plan, Migrate, Validate & Manage.
- Join our CEO, Kunal Agarwal and our CTO, Shivnath Babu as they discuss how Unravel helps organizations drive value from modern data applications in this bite-sized video: Drive Value from Modern Data Applications with Unravel.

Upcoming Events

- Join us at the Unravel Booth K2 at AWS Summit, New York, July 11.
- Unravel will bring its AI-powered DataOps platform to VB Transform in San Francisco (July 10 and 11). Stop by Unravel booth 2304 or RSVP for a session here. Don't miss our CEO, Kunal Agarwal on July 11, at 10:40am on the main stage as he talks about "Automatic performance management using AI Ops chatbot".

Community Highlights

- The Evolution of Analytics with Data: Data and Analytics have been one of the most commonly used words in the last decade or two. This post looks at how the field has evolved during the years, what roles in the market are currently evolving and how they are reshaping businesses.
- Pinterest has open sourced Singer, their logging agent that sends data to Apache Kafka for centralized logging. It provides both Java and Python libraries for both text and thrift format data. It also supports logging in Kubernetes via a side-car service. It includes a heartbeat system and produces audit records for centralized monitoring and alerting.
- Putting ML in production I: Using Apache Kafka in Python: This post illustrates how one could use a series of tools (mostly Kafka and MLFlow) to help productionising machine learning. It builds up on a simple scenario, similar to the real-world use cases.

Resources

- Learn more about Unravel.

- Online Product Demo
- Unravel Partners
- Unravel Product Releases and Documentation
- Unravel Datasheet
- More Unravel News