unravel

Unravel Newsletter: May 2019

Greetings from Unravel,

Today, Spark is the most widely used technologies for AI/ML and data pipelines and is considered to be the most popular open source project on the planet, with more than 1,000 contributors from 250+ organizations.

Oh, and by the way, Apache Spark just turned 10 years old!

To commemorate this milestone, we have a Spark-themed newsletter this month. We'll start by discussing Unravel's rich Spark tuning and troubleshooting capabilities and then continue with community highlights on this topic. Also included are links to recent webinars and upcoming events that you might find interesting.

Tune your Spark Applications using Unravel

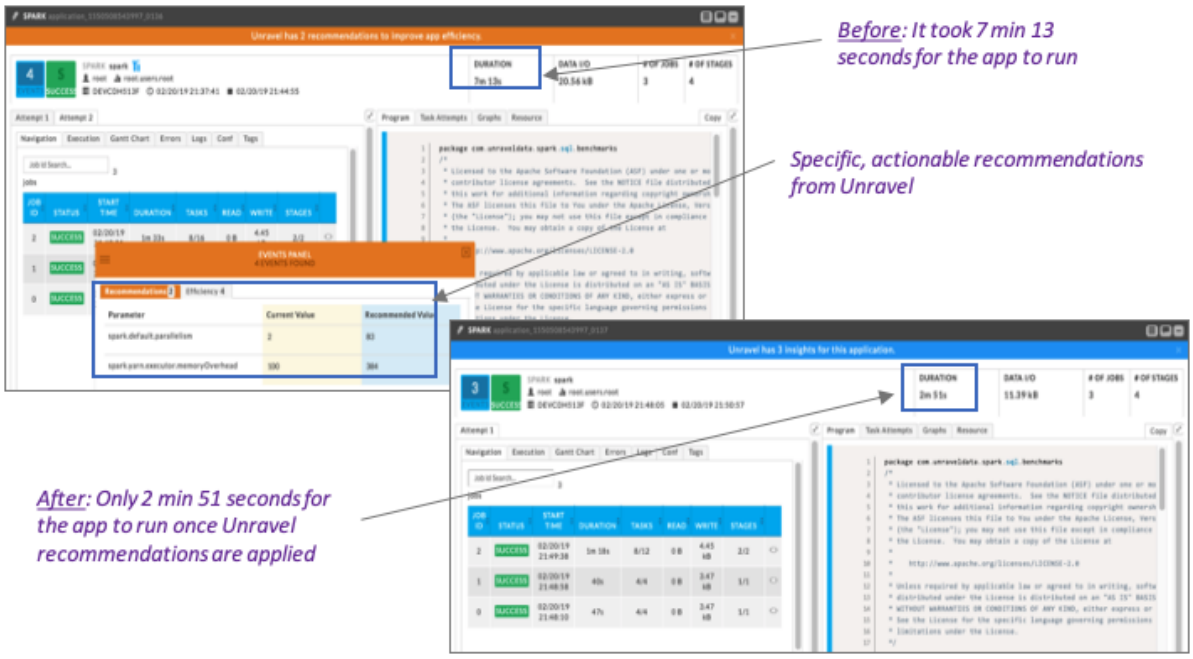
Spark applications are fairly easy to write and can be quite powerful when everything goes according to plan. But things get difficult when these apps start to slow down or fail. There could be a multitude of reasons - most of which are tedious and tough to track down and even tougher to solve.

Rishitesh Mishra, a Principal Engineer at Unravel, has been working on a blog series called **Why Your Spark Apps Slow Down or Fail**. Part I covers Memory Management and Part II covers Data Skew and Garbage Collection. Stay tuned for Part III.

Unravel automatically detects Spark issues like sub-optimal memory management and performance degradation from data skew, and many more.

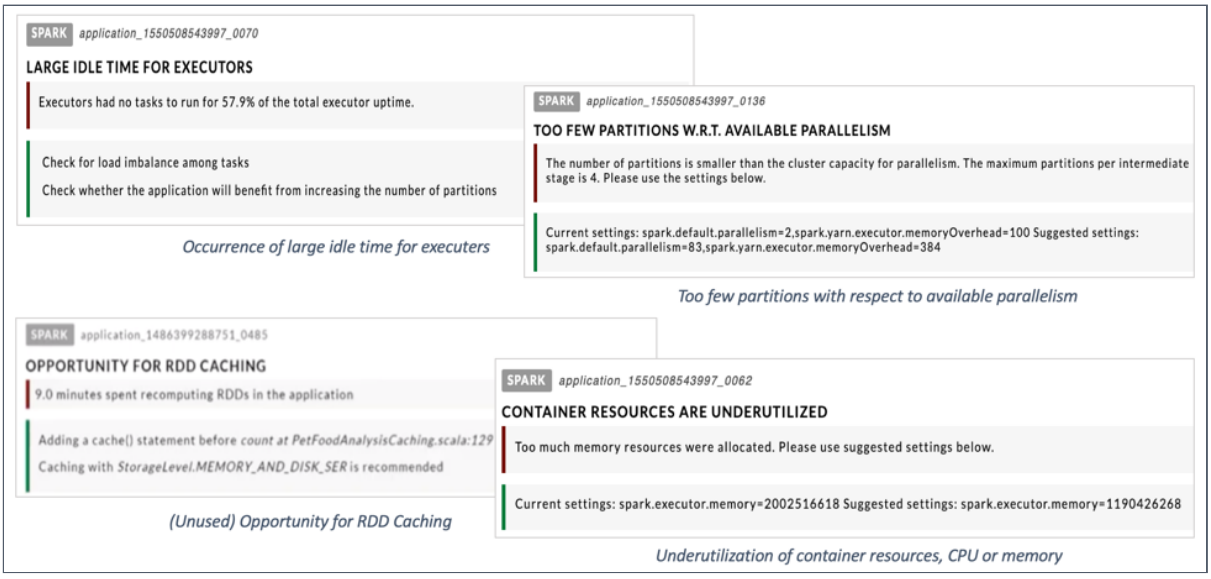
But Unravel goes beyond just monitoring and detecting problems. Unravel provides actionable insights and recommendations that help speed up your Spark applications and run them more reliably.

Here's an example. In the figure below, you can see that a Spark application takes over 7 minutes to run and that Unravel has provided some recommendations for changing some Spark tunable parameters. A subsequent run of the application took less than 3 minutes after changing these parameters.



Unravel Insights and Recommendations help speed-up your Spark Applications

These actionable and specific insights that Unravel provides cover a broad set of scenarios like the following:



Find more details about Unravel for Spark in the Unravel Solution Brief for Spark. Also, hear our CTO, Shivnath Babu talk about how Unravel uses AI to simplify Spark Performance and Operations Management: Putting AI to Work on Apache Spark. Read about how Unravel helps Reduce Apache Spark Troubleshooting time from Days to Seconds.

Don’t forget to check out the Spark Application Manager and Use Case - Optimizing the Performance of Spark Applications sections in the Unravel User Guide to help you get started!

Webinars

- Unravel Software Engineer, Alejandro Fernandez, explains how to assess, plan, execute, and validate a successful migration of data workloads to the cloud: Using AI-powered Automation for High Performance Data Pipelines in the Cloud

---

Conference Session Recordings

- Watch our talk from AWS Summit Santa Clara - Unravel: Migrating and Scaling Data Pipelines with AI on Amazon EMR, Redshift & Athena
- Watch our talk from Strata Data Conference San Francisco - Unravel: Automation of Root Cause Analysis for Big Data Stack Applications

---

Upcoming Events

- Join us at the Unravel Booth at AWS Summit, London, May 8.
- Join us at the #kafkasummit in London on May 13 at 4:15 pm where our CTO, Shivnath Babu will discuss how to use ML to identify root causes for a number of Kafka-based application bottlenecks, slowdowns, and failures. Details here.
- Join us at the Unravel Booth at AWS Summit, Stockholm, May 22.

Community Highlights

Keeping with the Spark theme this month

- Some well-regarded books to develop in-depth knowledge in Spark:
    - Learning Spark: Lightning-Fast Big Data Analysis: This book by Holden, Andy, and Patrick is one of the best Apache Spark books for starters as it discusses the Spark fundamentals and architecture. You will learn how to express parallel jobs with just a few lines of code, and cover applications from simple batch jobs to stream processing and machine learning.
    - High Performance Spark: Best Practices for Scaling and Optimizing Apache Spark: Aimed at people who already have an existing knowledge

of Apache Spark, this book discusses the best practices used in optimizing and scaling Apache Spark applications.

- Here is an interesting example of using PySpark to analyze httpd access logs and visualize the results with matplotlib.

Resources

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

Contact Us. Sign Up for 30-day Trial.