Unravel Newsletter: October 2019

Greetings from Unravel,

In this month's newsletter, we discuss two new reports (Executive KPIs Report and End User Report) that we added to Unravel's portfolio of On-Demand reports recently. The goal of both these reports is to make the most useful information for you available at your fingertips whenever you need it. Next, we discuss about the Unravel Customer Support process and share some useful articles. We will 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.

Like several of Unravel's features, these two new reports too are a direct outcome of us observing and hearing about needs from Unravel customers like yourself.

Unravel Executive KPIs Reports

This reports lets you quickly visualize the overall health and usage of your cluster. It contains the main KPIs of the cluster, including node health, apps and events over the chosen time period. You can visit the Unravel UI to generate it when you need it or schedule updates to be emailed to you periodically. In its current incarnation, this report can be used to:

- Present Executives with Bird's Eye View of Data Platform
  - High Level KPIs indicating Magnitude of Operations
  - Picture of Overall Health of Cluster
  - Rolled-up Consumption of Resources
- Get an overview of Utilization of Platform across Departments & Projects
  - Understand Value Provided and Cost (Chargeback) to User

A few sample sections from a report like this:

OVERALL HEALTH OF PLATFORM					
TOTAL APPS	SUCCESSFUL	FAILED	UNRAVEL EVENTS	READ I/O	WRITE I/O
168406	168348	6	1763	1.3TB	496.7GB

Overall health of the cluster - KPIs across the entire cluster
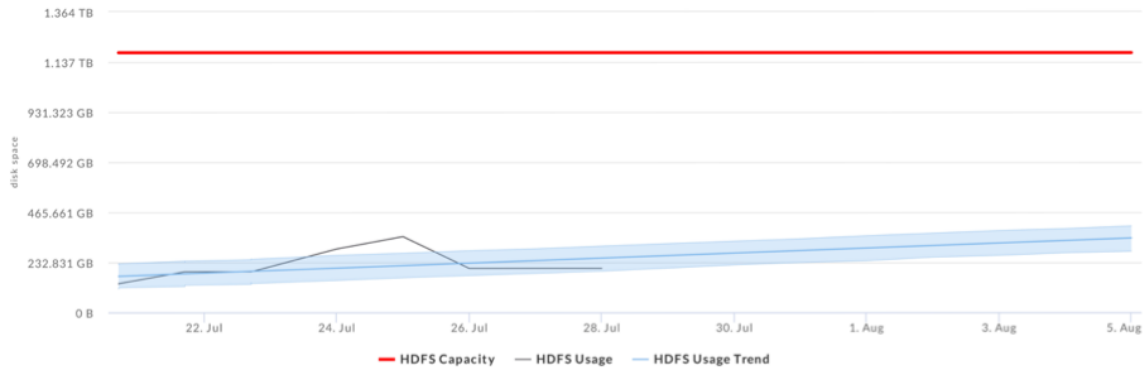
Resources: Allocated vs available - indicating if the cluster is over- or under-provisioned

HDFS forecasting Report

BY OUTPUTTABLES

OUTPUTTABLES	APP COUNT	CPU HOURS	% CPU HOURS	MEMORY HOURS	% MEMORY HOURS
[]	159	2.36	48.42%	5M 9D 7h 12m 45s	37.88%
[default.unravel_hdfs_fsi...	34	0.19	3.83%	20D 13h 37m 13s	4.89%
[default.unravel_hdfs_fsi...	14	0.21	4.37%	24D 7h 2m 25s	5.78%
[default.unravel_hdfs_fsi...	14	0.34	7.07%	1M 6D 23h 18m 37s	8.79%
NULL	182	1.77	36.30%	5M 29D 10h 10m 30s	42.66%

BY REALUSER

REALUSER	APP COUNT	CPU HOURS	% CPU HOURS	MEMORY HOURS	% MEMORY HOURS
unravel	209	2.98	61.36%	7M 23D 2h 36m 12s	55.43%
hive	8	0.04	0.82%	1D 19h 2m 10s	0.43%
NULL	186	1.84	37.81%	6M 5D 15h 43m 8s	44.15%

BY PROJECT

PROJECT	APP COUNT	CPU HOURS	% CPU HOURS	MEMORY HOURS	% MEMORY HOURS
group16	21	0.21	17.24%	21D 6m 33s	21.84%
group02	18	0.29	23.50%	20D 23h 36m 13s	21.81%
group05	16	0.26	21.29%	16D 21h 13s	17.54%
group15	17	0.17	13.69%	15D 19h 38m 41s	16.44%
group14	14	0.30	24.27%	21D 12h 21m 14s	22.37%

BY UNRAVEL.APP.NAME

UNRAVEL.APP.NAME	APP COUNT	CPU HOURS	% CPU HOURS	MEMORY HOURS	% MEMORY HOURS
SELECT parent_path, c...	60	0.48	30.45%	1M 17D 5h 23m 30s	29.15%
SELECT parent_path, c...	59	0.38	23.85%	1M 3D 23h 34m 54s	20.98%
INSERT INTO TABLE unr...v...	17	0.10	6.04%	10D 12h 56m 51s	6.51%
INSERT INTO TABLE unr...d...	15	0.28	17.63%	1M 13h 43m 45s	18.87%
INSERT INTO TABLE unr...l...	15	0.35	22.03%	1M 9D 16h 5m 31s	24.49%

BY USER

USER	APP COUNT	CPU HOURS	% CPU HOURS	MEMORY HOURS	% MEMORY HOURS
hive	379	3.58	73.67%	1Y 1D 22h 58m 30s	86.07%
unravel	24	1.28	26.33%	1M 28D 14h 23m 2s	13.93%

YARN (and Impala) Consumers - by Dept, Project, Realuser, App Name etc.

Check out this example report sent to a user. Also see this section in the Unravel User Guide to help you get started with using this report (also referred to as. The "Cluster KPIs Report").

Keep looking at the newsletters for upcoming exciting additions to this report or get in touch with us directly to get your feature requests in!

Unravel End User Reports

This report is designed keeping the following goals in mind:

- Help end users execute their jobs more effectively
- View the Performance & Resource Usage of Apps
- Get a quick and clear overview of how your apps are performing on the platform, how much resources they are taking up in running their apps and recommendations/scope for improvement.

You can visit the Unravel UI to generate it when you need it or schedule updates to be emailed to specific users periodically.

This report can help you determine what apps need attention to improve duration, resource usage, etc. Also, if you are a Team leader or manager, you can use it to track how the team uses resources and identify users who overuse resources (rogue users). You can get a concise and clear view of:

- How apps are performing on the platform
- The resources the apps are using
- The number of recommendations or insights Unravel has for the app.

"Spark" Section of the User Report

Check out this example report sent to a user. Also see this section in the Unravel User Guide to help you get started with using this report (also referred to as the "User Report").

Unravel Customer Support

Many of you are accustomed to logging support requests via the Unravel Customer Support Portal. We have enjoyed working with you on those and hope that you have

had a good experience as well. We are continuously evolving and improving our processes, tools and knowledge base to ensure you get the best experience (both in terms of quality and timeline of resolution of issues). Please continue to use the portal to:

- Log a Case to obtain assistance from Unravel Customer Support
- Search our Knowledge Base of How to Questions and Technical Issues
- Create and maintain Environment records for each of your Unravel Environments
- View all of the feature requests logged by your company

Our knowledge base now spans over 100 KB Articles. Here are 3 recent ones that you may find useful:

- Spark Applications submitted through Oozie do not show metrics in Unravel Application UI
- How to Check for Processing Lag within Unravel?
- How can I debug Unravel Kafka Lag in terms of time?

Please reach out to us for any feedback and suggestions for our overall support and engagement process. We look forward to hearing from you.

Conference Sessions & Webinar Recordings

- Watch this webinar by Alejandro Fernandez, Principal Software Engineer, Unravel - Using AI powered Automation for High Performance Data Pipelines in the Cloud.
- Watch this demo by Abha Jain, Director, Products, Unravel - Unravel for Azure Databricks overview demo.
- Check out this video Unravel and Clearsense chief executive discuss the potential life and death challenges of big data in healthcare - Fireside chat with Clearsense CIO Charles Boicey and Unravel CEO Kunal Agarwal.
- Watch this webinar by Abha Jain, Director of Products, Unravel Data, and Ron Abellera, Microsoft Global Blackbelt Microsoft - APM & Operational Intelligence for Azure Databricks

---

Upcoming Events & Webinars

- Join us at the Unravel Booth at AWS Summit, in Toronto on Oct 3.
- Join us at the Unravel Booth at Spark AI Summit EU, in Amsterdam, Oct 15 – Oct 17
- Looking forward to seeing you at the Phoenix Data Conference, in Phoenix on Nov 09.

Community Highlights

- **An Engineer's Perspective on Engineering and Data Science Collaboration for Data Products**: At Coursera, engineers and data scientists have built many data products. They've learned that building a data product is a team sport. This post outlines three themes that worked well in their pursuit of this goal from an engineering perspective.
- **How to Unlock the Full Potential of Kafka Producers**: Gojek shares some tips for configuring and tuning the Kafka Producer
- **Explaining SQL and NoSQL, to Grandma**: Over the last 15 years, many new databases have come to the market as part of the No-SQL movement. These include key-value stores such as Redis and Amazon DynamoDB, wide-column stores such as Cassandra and HBase, document stores such as MongoDB and Couchbase, and graph databases and search engines such as Elasticsearch and Solr. This is a high-level overview of SQL and NoSQL.
- **Are you sure you're using microservices?**: A good list of best practices in a microservices architecture. If you're feeling pain with your deployments or architecture, then this post has lots of ideas for how to improve (like making sure you can upgrade a database schema without updating multiple services). The author notes that a half-baked microservices architecture can lead to failure scenarios that are more likely than they would be in a monolith.
- **7 mistakes when using Apache Cassandra**: A good list of pitfalls and anti-patterns both for those getting started with and looking at expanding usage of Apache Cassandra. For example—it's important to know your query pattern before you model your data, and Cassandra shouldn't be used as a queue. The post covers seven potential mistakes in some detail and mentions a few others to keep an eye out for, too.

Resources

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