

Unravel Newsletter: December 2019

In this month's newsletter, we continue to showcase Unravel capabilities for migrating big data workloads to the cloud that we had discussed a few months back. We also introduce you to the Unravel Cluster Compare Report. 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.

Looking to Migrate your Big Data Workloads to the Cloud?

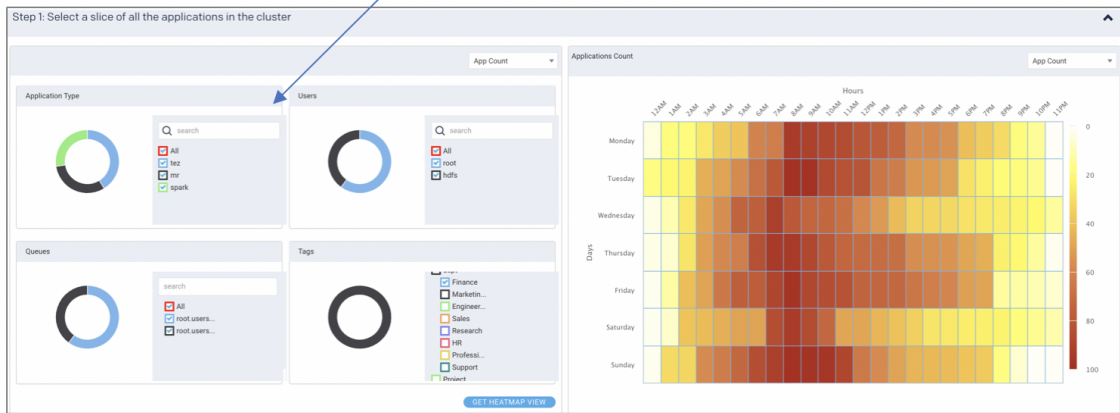
A few months ago, we mentioned Unravel's intelligent migration planning and assessment capabilities help solve various challenges related to migrating big data workloads to the cloud:

- Adequate understanding of the current environment
- Prioritization of applications to move to cloud
- Optimal Cloud Topology Design
- Cost estimation
- Performance that meets or exceeds the source on-prem cluster

Specifically regarding "Optimal Cloud Topology Design" and "Cost Estimation", we had discussed how you can generate an Unravel report that would provide you with the optimal mapping of your on-premise cluster to a deployment topology in the cloud. This is based on your goal/strategy for the migration (minimizing risk/reducing cost) and resulting cost. Now, we are adding an exciting new feature to the existing portfolio of migration planning and assessment features where you can obtain the optimal cloud topology design (and cost estimate) corresponding to a slice of applications running on your cluster. Here's a quick description of how this would work:

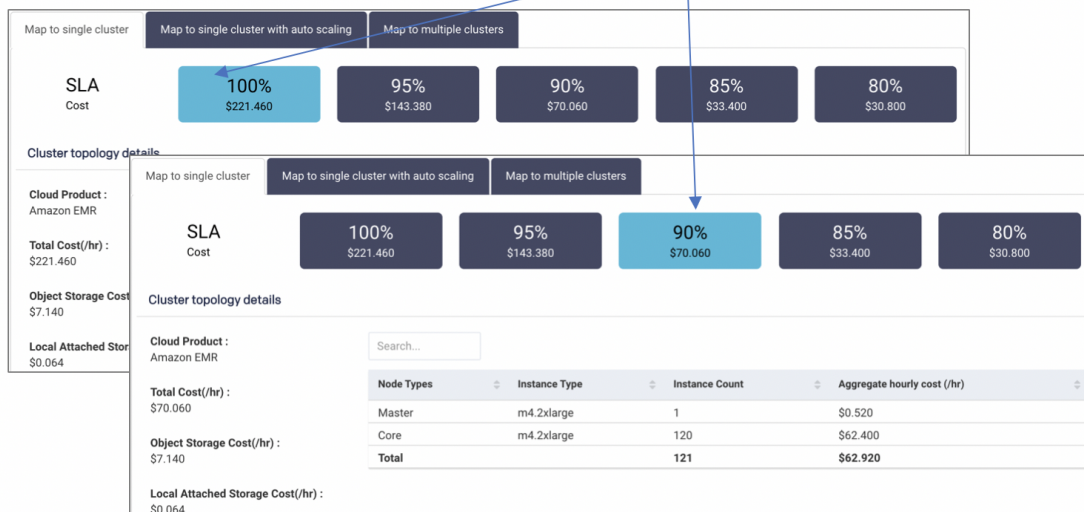
- First you would select the desired slice of applications from your on-prem cluster that you would like to migrate to the cloud:

Select the desired slice of applications from the on-prem cluster



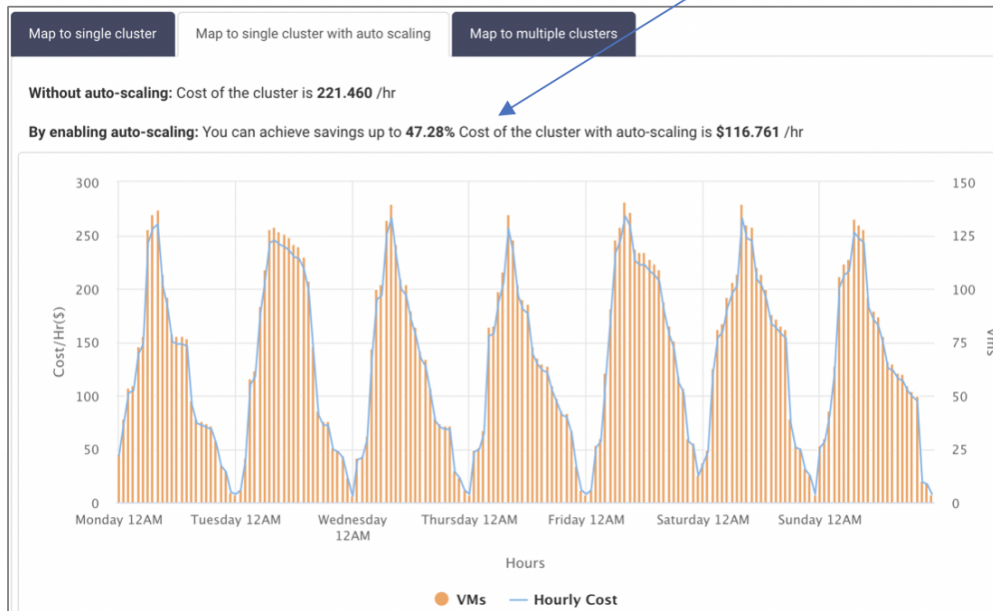
- Once you have chosen the desired target platform in the cloud and its service (e.g. Amazon EMR) and run the “Workload Fit” report, you can see results as shown in the following sample:
 - “Map to single Cluster”: You get the recommended cloud cluster topology for the slice of applications you selected. In addition, you get an option to exercise price-performance trade offs. For example, when you click on 100% for SLA, you get the estimated cost and topology of the cluster if you wanted the apps to meet their resource requirements a 100% of the time. However, if you are okay with some SLA degradation to achieve cost savings, we show you options for that as well.

Recommended Cloud Cluster Topology with Cost Estimates
(including ability to exercise price-performance tradeoff)



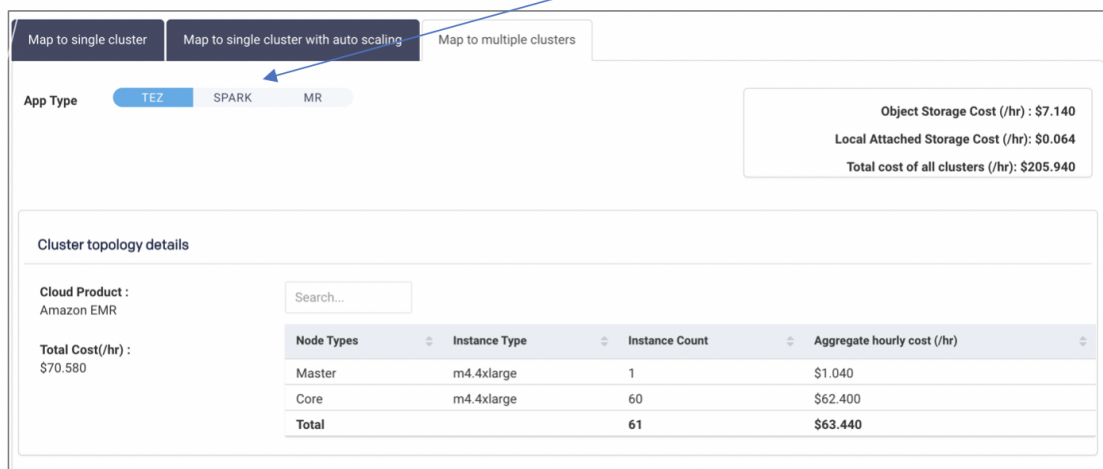
- “Map to single cluster with auto scaling” - We also tell you how much you potentially can save if you were to configure auto scaling in your cloud cluster. This data once again, is based on actual data from your cluster and hence scientific and real!

Alleviate fear of migrating to cloud by observing evidence that cloud-native features can reduce costs (based on actual data from your on-prem environment)



- “Map to multiple clusters” - You can also choose to migrate your selected slice of applications to multiple clusters - each running applications which are of different types. This allows you to potentially set appropriate cluster optimization parameters.

Split your chosen slice of apps into three different clusters in the cloud: One for Tez, one for Spark and one for MR jobs



Unravel Cluster Compare Report

This report helps in comparing the cluster resource utilization for two chosen periods of time and give an option to group by user or queue. Such reports can help you detect and visualize deviations from the norm/expected level of resource usage. Here, some example screenshots of this report are shared:

GROUP BY

User

Queue

Time Range:

Last 90 Days

Compare With Range:

Last 30 Days

QUEUE	VCORE				MEMORY (GB)				RUNNING APPS				PENDING APPS			
	MAX	MIN	AVG	STD DEV	MAX	MIN	AVG	STD DEV	MAX	MIN	AVG	STD DEV	MAX	MIN	AVG	STD DEV
root.users.hdfs																
Sep 1, 2019-Nov 29, 2019	47	0	5.27	7.67	56.25	0	15.40	9.18	2	0	0.88	0.39	1	0	0.25	0.43
Oct 30, 2019-Nov 29, 2019	47	0	6.13	7.52	56.25	0	16.01	9.31	2	0	0.94	0.35	1	0	0.24	0.43
root.users.root																
Sep 1, 2019-Nov 29, 2019	48	0	10.09	6.98	90	0	47.63	19.31	5	0	1.01	0.24	1	0	0.02	0.14
Oct 30, 2019-Nov 29, 2019	43	0	5.53	4.56	90	0	30.05	17.72	5	0	1.11	0.47	1	0	0.04	0.18

Cluster compare – Group by Queue

GROUP BY

User

Queue

Time Range:

Last 90 Days

 Compare With Range:

Last 30 Days

	VCORE				MEMORY (GB)				RUNNING APPS				PENDING APPS			
USER	MAX	MIN	AVG	STD DEV	MAX	MIN	AVG	STD DEV	MAX	MIN	AVG	STD DEV	MAX	MIN	AVG	STD DEV
root																
Sep 2, 2019-Nov 30, 2019	48	0	10.09	6.98	90	0	47.63	19.31	5	0	1.01	0.24	1	0	0.02	0.14
Oct 31, 2019-Nov 30, 2019	43	0	5.53	4.56	90	0	30.05	17.72	5	0	1.11	0.47	1	0	0.04	0.18
hdfs																
Sep 2, 2019-Nov 30, 2019	47	0	5.27	7.67	56.25	0	15.40	9.18	2	0	0.88	0.39	1	0	0.25	0.43
Oct 31, 2019-Nov 30, 2019	47	0	6.13	7.52	56.25	0	16.01	9.31	2	0	0.94	0.35	1	0	0.24	0.43

Cluster compare – Group by User

Check out [Section: Cluster Compare](#) in the Unravel User Guide to help you get started with using this report.

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

Conference Sessions & Recordings

- Watch this video for a demo of Unravel Migration Planning and Assessment capabilities by Abha Jain, Sr. Director, Products, Unravel - [Migrating Big Data Workloads to the Cloud with Unravel](#)
- Abha Jain also provides a demo of Unravel data insights including KPIs for tables/partitions, data tiering et al - [Unravel Data Insights Demo](#)

Upcoming Events & Webinars

- Join us as we discuss Unravel for Amazon EMR with Shashi Raina, AWS cloud architect: [Always Meet Your SLAs on Amazon EMR](#).
- Join us at the Unravel Booth at [Strata Data Conference San Jose 2020](#), in San Jose, March 16 – March 18, 2020.

Community Highlights

- [Five companies making humans and AI collaborate](#): Nice post on how companies like Unravel are using AI to bring innovative solutions for hard problems.
- [Open Sourcing Amundsen: A Data Discovery And Metadata Platform](#): Lyft writes about Amundsen, their open-source tool for metadata management and data discovery. The post covers the architecture of the system—the metadata, search, and frontend services as well as the databuilder (data ingestion framework). Since open sourcing earlier this year, they've had a number of contributions, such as one to make the datastore more extensible (supporting Apache Atlas in addition to Neo4j).
- [GraphQL Search Indexing](#): Netflix describes how they use GraphQL to build indexes from data stored across multiple services. Their key idea is to do issue a (batch) GraphQL query to return a full denormalized record (e.g. a show and its episodes, etc.) and store the results in Elasticsearch. Next, they can listen to changes on Kafka, and follow the relationships from the GraphQL schema to invalidate/reindex data.
- [Spark Job Optimization: Dealing with Data Skew](#): A look at how to detect skew in your Apache Spark jobs, and several ways to fix a job with skew (hints, randomizing the join key, writing a custom partitioner). Which solution is best/fastest depends a bit on the inputs to your job.
- [Redshift's Automatic WLM with Query Priority: A First Look at Performance - intermix.io](#): In this article, Intermix is looking at the performance improvements by the new feature with Automatic WLM that we're really excited about: Query Priorities.

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](#). [Take Unravel for a Test Drive](#).

2019 Unravel. All Rights Reserved. 2 Palo Alto Sq, Suite 120, Palo Alto, CA 94306