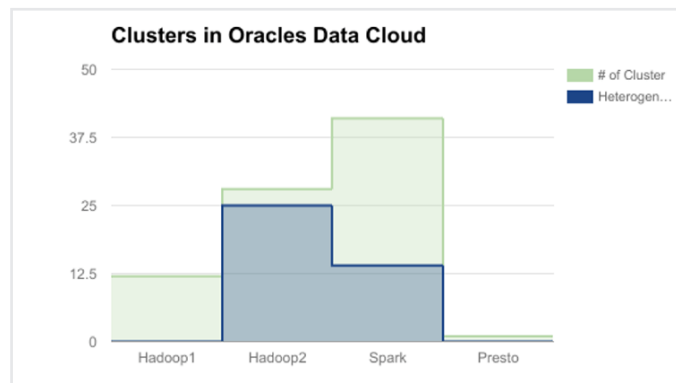


Oracle uses Heterogeneous Cluster to achieve cost effectiveness

About Oracle Data Cloud

Oracle Data Cloud has 82 clusters with Qubole, the distribution and heterogeneous usage are as following:



Cluster Type	Total # of Cluster	Configured with Heterogeneous
Hadoop1	12	0 (Not Supported)
Hadoop2(Hive)	28	25
Spark	41	14
Presto	1	0

“ As our EC2 costs kept climbing and the spot market became more volatile, heterogeneous was the only option that made sense.

Even as our usage has grown over the past 6 months, since switching to heterogeneous, our costs have either gone down or, at least, stayed the same.



JUSTIN WAINWRIGHT
System Analyst,
Oracle Data Cloud

Challenges

At Oracle Data Cloud (ODC), big data is our business. We tend to use larger clusters and process for sustained periods of time. Over the past couple years, we've seen the demand for certain instance families increase and put the health of our clusters at risk. As our EC2 bills were already climbing at an alarming rate, we couldn't justify moving to more on-demand nodes, so we needed to keep using spot nodes in the majority of cases while keeping clusters stable.

"Because we run almost 100% spot nodes, we were suffering catastrophic spot losses given the size of our clusters and the scale at Oracle," explains Justin Wainwright, System Analyst Oracle. "This resulted in missing SLA, job failures and most importantly wasted time on business pipeline. Heterogeneous was the feature we were looking for."

Oracle Data Cloud Team became the first beta customer for heterogeneous cluster feature when Qubole launched it back in August 2016. They were excited about the great potential this feature can bring to their daily operation.

Justin and his team started with smaller, non-critical operation clusters to test the water then with positive results. They expanded the configurations to entire cluster fleet Oracle Data Cloud owns. Qubole has been doing cost comparison and analysis along with Oracle for this amazing transformation journey – based on the statistics, we've seen up to 90% cost saving compared to on-demand node cost and usually 20-50% cost saving compared to homogeneous spot configurations.

Oracle Data Cloud team also helped Qubole to make a better product during this journey and shared their experience in configuring heterogeneous cluster with other Qubole users, such as:

- For long-running jobs: split jobs into smaller chunks to better adapt to smaller heterogeneous instance types (e.g. make size appropriately for executors and overhead settings). Make the size based on average workload
- For burst jobs: choose stronger heterogeneous instance types (e.g. 10xLarge against default 4xLarge)
- For Outliers: you need custom configurations to achieve the best heterogeneous cluster performance – this would require some trials and tweaks over time.

Results

Oracle Data Cloud was able to reduce spot loss and significantly lower the total operation cost further with heterogeneous configuration.

At pre-heterogeneous peak usage (Fall 2016), almost 40% of their EC2 costs were Qubole jobs running on on-demand nodes. As of May 2017, on-demand costs are no more than 20% and nearly half of the clusters have been configured as heterogeneous.

Cost Comparison:

- September 2016: 34K USD spent on-demand nodes
- January 2017: 17K USD spent on-demand nodes

QDS empowers Oracle Data Cloud operation team with the capability to scale out fast while keeping the cost low and their customer satisfaction high.

What's Next?

Oracle and Heterogenous Clusters – Justin Wainwright, Oracle

About Qubole

Qubole is passionate about making data-driven insights easily accessible to anyone. Qubole customers currently process nearly an exabyte of data every month, making us the leading cloud-agnostic big-data-as-a-service provider. Customers have chosen Qubole because we created the industry's first autonomous data platform. This cloud-based data platform self-manages, self-optimizes and learns to improve automatically and as a result delivers unbeatable agility, flexibility, and TCO. Qubole customers focus on their data, not their data platform.