



Industry: Media

How MiQ Partners with Qubole to Drive Sustainable Growth for their Consumer Data-Driven Business

In mid-2017, programmatic media partner MiQ could see that their data lake ecosystem needed new capabilities to scale at par with the company's growth. They realized they needed to make changes if they were to scale cost-effectively to their own growing data needs and those of their clients.

Up until that time, MiQ had been using AWS EMR as their data platform and Redshift as their cloud data warehouse. But they were having difficulty meeting their SLAs to their user community.

"We were running into scaling issues with EMR," says Rohit Srivastava, Engineering Manager at MiQ. "Our requirements were growing, but EMR wasn't scaling. That's when we began discussions with Qubole. We had not seen a platform which offered such a rich array of technology support."

Searching for a scalable platform

Being a data-driven business, MiQ maintains an extensive data infrastructure. They were running AWS EMR, Apache Spark, Presto, Apache Hive, AWS Redshift, AWS Athena and Apache Airflow. They had shuttling capabilities and a variety of other tools. What they needed was a platform that was both scalable and responsive.

Cost was a major driver. The platform had to be economical. From 2015 to 2018 the cost of the existing platform based on EMR and Redshift had risen sharply and was becoming unsustainable. So infrastructure and revenue growth had to find balance.

"Our entire business model is data-driven," says Rohit Srivastava. "Because of this, we were very mindful of what we were spending on and what kind of ROI we needed to achieve."

MiQ decided to run a Proof of Concept—with Qubole's assistance—optimizing their four most complicated use cases on the Qubole platform across varied workloads of Hive, Spark and Presto. Within six months, Qubole had proven itself—both in terms of cost-effective scaling and in terms of allowing MiQ to meet its SLAs—for all four use cases.

MiQ signed a contract with Qubole and began full-scale onboarding of the platform in November 2018.

About MiQ

MiQ is a leading programmatic media partner for brands and agencies. Headquartered in London, MiQ has offices across North America, Europe and Asia Pacific. We work with the world's leading brands and media agencies such as Marriott, Dell Mercedes, Microsoft, GroupM, Dentsu Aegis and IPG. We were named 4th in The Sunday Times International Track 200 for 2019, the Fastest Growing Tech Company of the Year at the 2017 Stevie Awards and awarded Most Effective Use of Data at The Drum's Digital Trading Awards USA 2017. MiQ operates globally from 18 offices located in North America, Europe and APAC.

Better performance on visualizations and insights... at lower cost

Among MiQ's most complicated use cases are those related to its Campaign Optimization Dashboard. They had been running on a legacy data warehouse system based on Redshift and an interactive Spark cluster, supported by a data cube of roughly 10 to 20 GB of data per day. MiQ wanted to give users a month of look-back—an operation requiring terabytes of data.

But there were two main problems with this Campaign Optimization Dashboard.

The first was cost. MiQ needed to run the dashboards on an interactive cluster. But running an always-on Spark cluster with their existing infrastructure was costing them a considerable amount of money, which would soon become unsustainable.

“

For us, visualization and insights are key. Presto on Qubole gave us superior numbers on one of our costliest and heaviest dashboards. That was very motivating, and it became a deciding factor for us to move ahead with Qubole.”

Rohit Srivastava

Engineering Manager, MiQ

The second problem was performance. MiQ needed rapid turnaround to meet their SLA for this service. MiQ's Data Engineering team worked hard to optimize Spark on its existing infrastructure for these dashboards, but achieving the necessary performance was challenging.

As part of their Qubole proof of concept, the Data Engineering team at MiQ ran a trial of Operational Dashboards on Qubole/Presto against Spark. After two months of thorough performance benchmarking against multiple workloads, Presto on Qubole emerged as the clear winner, not only on performance but also on cost.

“Doing the same thing in our existing Spark-based stack would have cost us two and a half to three times more than what we saw with Presto,” says Rohit Srivastava. “So, switching our optimization dashboard to Presto on Qubole clearly made sense.”

Over one hundred happy business analysts onboarded in just one month

MiQ's analytical capabilities are consumed and best translated to their clients success by a group of highly skilled and efficient Business Analytics specialists. Interfacing directly with MiQ's customers, the Business Analysts, work on some very tight SLAs, often generating and delivering insights every week or even every three days.

Rohit cites a typical example of running a Hive query over a terabyte data set. If the business analysts are running into delays due to poor platform response times, and they have to spend a couple of days gathering the results from that query, they end up with a lot less time to prepare their insights report for the customer.

A key business OKR (objective and key result) for MiQ is that only 20% of a business analyst's time should be spent in data preparation and gathering, while 80% should go into creative analysis to figure out what insights can be delivered to customers. A key KPI for meeting this objective is response time to data queries.

MiQ found they needed to improve platform query response times by 300 to 500 percent if their business analysts were to meet their SLAs. ETL times needed to drop from hours to minutes. Thanks to results obtained previously, MiQ knew they could achieve that performance with Hive and Spark on Qubole. But what amazed them was the speed at which they were able to get their analysts up and running with Qubole. Initial adoption metrics indicated the transition might take two to three months. Instead, 100%—north of 100 Business Analysts—were onboarded within one month.

“For the Business Analytics team, it was a happy and smooth shift,” says Rohit Srivastava. “They were on EMR, and EMR’s interactive mode is challenging to use. The moment they experienced Qubole—be it in terms of performance, result logs, not waiting long for clusters to come up, downloading results, scheduling their workflows—the end user experience for those business analysts improved greatly.”

Containing interactive data science costs

Another stakeholder in MiQ’s data lake infrastructure is its data science team. This team of data scientists had been running queries on Spark since 2014 on EMR.

“Running Spark queries on EMR was not optimal in terms of infrastructure, cost or performance,” says Rohit Srivastava.

MiQ’s data scientists need to do most of their work in interactive mode. They are building, testing, training, running and monitoring models. They’re running queries and building analyses, and their interactive usage is extremely high. MiQ’s data science team was well aware of the pain points of EMR, like the lack of autoscaling and clusters going down frequently. So, when they were introduced to Qubole and Spark, they were happy to climb aboard. Much of the team’s work involves building, testing and monitoring of models, as MiQ’s entire business depends upon predictive analysis.

MiQ deploys numerous models at any given time and each provides a strong look-back ranging over three to six months of data. These models are constantly running to deliver predictions, ranging in frequency from a day up to a week. In some scenarios they may run predictions every few hours as well as in near real time. “Today, all our interactive usage and the majority of the products are being developed on Qubole” says Rohit Srivastava.

Rohit explained that switching to Qubole also resulted in huge cost reductions. Since making the transition, they are saving:

- 200% improvement in monthly interactive cluster charges
- Significant 5-figure annual savings through application of weekend scripts
- 20% reduction of monthly costs by using simplified cluster configurations

Qubole makes this third saving possible by running Presto on a single cluster, referred to as “Unified Insights Platform” within MiQ, without impacting SLAs—a more economical solution than the multiple Spark clusters or the Redshift data warehouse solution. Other savings have been realized thanks to ease of use, ease of maintenance, and by having 90% of applications running on a single point of entry for Presto.

Easy application development

Supporting its business analysts and data scientists, MiQ also has developers manning several tech teams. One of the teams is the Data Processing team which is constantly building scalable microservices for both Data Engineering and Data Science modules on top of Qubole and MiQ’s other data platforms.

Rohit says MiQ has found it challenging to develop applications and services on top of some other platforms, because those don’t offer a rich application programming interface (API). In contrast, Qubole’s API ecosystem makes application development and integration easy.

Rohit and his colleagues have been pleased and proud of the collaboration between MiQ and Qubole in developing applications for their users. “Developers from the Data Processing team at MiQ even made open-source contributions which Qubole has built into some of its APIs,” says Rohit Srivastava. “And in several cases, where we needed APIs quickly, the Qubole team was really agile in building those APIs for us. That was something which was truly outstanding.”

How MiQ measures Qubole’s value to their business

MiQ uses numerous KPIs to measure the cost-effectiveness of their data lake and the various platforms and tools they use to exploit and manage their data. Rohit groups these in three domains: stability, cost and skillset.

Greater pipeline stability

A key performance indicator of stability is pipeline failures. Failures of automated workflows built by business analysts and data scientists—how many pipelines failed and were not able to be utilized by the downstream or upstream jobs—are tracked on a weekly and monthly basis.

For example, in an MiQ application called “Analytics Platform”, workflows often integrate code written in Python and R. These pipelines used to fail frequently. One study MiQ performed showed 15% to 20% of their pipeline failures were caused by Python and R instability, due to these running within the application infrastructure in a non-distributed architecture.

“

What’s really awesome about Qubole is that for anything that exists in a Qubole user interface, you have an API built for it. It’s easy to program an integration. Plus, Qubole’s team has been very helpful there, which has been good for us.”

Rohit Srivastava

Engineering Manager, MiQ

Then Qubole entered into the picture. Now, instead of running Python and R within the Analytics Platform itself, MiQ has integrated them through the Qubole API. By running the entire Python and R-indigenous stack on Qubole, they have brought that 15% to 20% failure rate down to just 1-2%.

Savings across the board... and better cost management

One of MiQ’s primary OKRs states that the rate of infrastructure growth must not exceed 65% of the revenue growth rate. Rohit says the Data Engineering team at MiQ was supported by Qubole to meet that OKR in several ways.

In addition to the aforementioned significant annual saving realized through the Spark to Presto on Qubole migration, Qubole has also helped MiQ lower costs through optimization of its AWS S3 object storage, data cleanup, optimization of instance types, and reduction of nodes.

Qubole even helps with cost monitoring and management. MiQ has built a rich cost reporting microservice using Qubole’s and AWS’s Cost Explorer, which interfaces with most of the data technologies MiQ uses. This microservice collects and summarizes data and automatically distributes Cost Explorer reports, giving upper management weekly and monthly snapshots of performance

against the OKR. It also sends daily deviation reports to line managers, so they can alert their teams and adjust immediately. MiQ's Tech will team were open-sourcing this service as a generic solution for Cost Management very soon.

Building employee skill set

Rohit says the optimization work MiQ has done since adopting Qubole has been instrumental in improving the skill set of their team. "We have been able to do a lot of knowledge transfer to our Business Analytics team and Data Science team through Hive and Spark optimization, Presto for the Business Intelligence team and so forth," he says. Before Qubole there was no concept of a shared infrastructure, but a fragmented approach. Now all teams work with data engineers on optimizations and knowledge sharing, as opposed to siloes. This has also helped the Data Engineering team at MiQ scale in terms of multiple frameworks, technologies and languages that the group is equipped with today.

Even the way MiQ has approached cost savings with Qubole has been valuable in skillset building, not only in terms of understanding, but also in terms of responsibility and accountability. "I think accountability is one of the key areas," says Rohit Srivastava. "If we want to sustain our revenue and our organizational and employee personal development, accountability plays a very key role."

A vision for the future

What are MiQ's immediate and long-term plans for Qubole? Rohit divides their plans into two categories.

The first of these is sustainability. "We need to make sure we sustain and build upon what we have already built," he says, "Be it in terms of stability, cost or skill set, sustainability is something we want to track continuously. Qubole really helps with that."

The second area is creativity and innovation. Key areas of innovation for the near-term include privacy, migration of MiQ's Athena accounts to Presto, and data real-time streaming.

Ensuring consumer privacy

Privacy is a key theme for MiQ. Their business is driven by how consumers think and by the online and offline activity of those consumers. Thus, like most businesses, they have ongoing consumer privacy initiatives being driven by GDPR, CCPA and other legislation.

One such initiative is a project in MiQ called Data Minimization. This is an effort to monitor all incoming data feeds and eliminate or obfuscate all personal identifiable data in those feeds to protect consumer privacy. The Data Ingestion team at MiQ is building the minimisation solution where the Qubole plays a key role as a processing layer. The project covers some 40 to 50 different big data pipelines running on Qubole in the most cost effective mode. Rohit says it is one of MiQ's key focus areas and deliverables for early 2020.

Migrating from Athena to Qubole Presto for better value and performance

MiQ traditionally has had a decent AWS Athena infrastructure for making queries and analyzing data. Since Athena charges by the query, however, this solution was becoming less and less cost-effective as MiQ grew.

Fortunately, Presto on Qubole eliminates those problems. Combine that with Qubole's workload-aware autoscaling, automated real-time Spot buying, and Presto on Qubole becomes a much more advantageous solution for MiQ.

"Working with Qubole, we are planning to move all our Athena usage to Presto this year, while still making sure we meet our SLAs," says Rohit Srivastava. "Athena-to-Presto migration is something we're really excited about."

A quest for real-time consumer retargeting

Another area where MiQ is using Qubole is in data streaming.

MiQ has been developing a series of what they refer to as “Predictive Retargeting Solutions.” These products are designed to gather data from target users’ on on-going events, generate insights, and enable users to make decisions and take action—all in real-time. In short, they are on a quest for real-time consumer retargeting leveraging Qubole Pipelines Service, for building and managing these streaming data pipelines.

One of the solutions, called CastIQ, is a political reporting dashboard aimed at the current political season in the United States. With CastIQ, users can query political data according to issues, trends, regions, candidates, etc., and gain insights based on up-to-the-minute data. It demanded very high performance and response time SLAs, and Presto on Qubole helped scale this solution with a quick turnaround time. At the time of this writing, CastIQ has already brought in significant revenue in less than nine months.

“Qubole support has been instrumental in allowing us to meet our SLAs for this project. MiQ and Qubole have worked really, really closely on this, and both teams did outstanding work.”

Rohit Srivastava

Engineering Manager, MiQ

Qubole is the open data lake company that provides a simple and secure data lake platform for machine learning, streaming, and ad-hoc analytics. No other platform provides the openness and data workload flexibility of Qubole while radically accelerating data lake adoption, reducing time to value, and lowering cloud data lake costs by 50 percent. Qubole's Platform provides end-to-end data lake services such as cloud infrastructure management, data management, continuous data engineering, analytics, and machine learning with near-zero administration. Qubole is trusted by leading brands such as Expedia, Disney, Oracle, Gannett and Adobe to spur innovation and to transform their businesses for the era of big data. For more information, visit us [online](#).

For more information:

Contact:
sales@qubole.com

Try Qubole for Free:
[Start 30 Day Free Trial](#)

469 El Camino Real, Suite 205
Santa Clara, CA 95050
(855) 423-6674 | info@qubole.com

WWW.QUBOLE.COM