Case Study | Qubole

The Big Data SaaS Company

A case study on how MediaMath uses Qubole, a Big Data Platform as a Service offering built on top of Apache, to solve Big Data problems in the world of Digital Ad-Tech.

"We needed something that was reliable and easy to learn, setup, use and put into production without the risk and high expectations that comes with committing millions of dollars in upfront investment."

Marc Rosen,

Sr. Director of Data Analytics

About MediaMath

MediaMath, a 260-employee company based out of New York City, founded in 2007, is the leading global digital media-buying platform. MediaMath develops and sells tools for Digital Marketing Managers under the TerminalOne brand. TerminalOne allows Marketing Managers to plan, execute, optimize, and analyze marketing programs. This is a case study written by MediaMath for Qubole.

Background

The Analytics and Insights team at MediaMath is responsible for delivering decision-making infrastructure and advisory services to our clients. The team does this by helping clients answer complex business questions using analytics that produce actionable insights. Examples of the team’s work includes but is not limited to:

- Segmenting audiences based on their behavior including such topics as user pathway and multi-dimensional recency analysis
- Building customer profiles (both uni/multivariate) across thousands of first party (i.e., client CRM files) and third party (i.e., demographic) segments
- Simplified attribution insights showing the effects of upper funnel prospecting on lower funnel remarketing media strategies

The Challenge

Our flagship product captures all kinds of data that is generated when our customers run digital marketing campaigns on TerminalOne. This data amounts to a few terabytes of structured and semi-structured data in a day. It consists of information on marketing plans, ad campaigns, ad impressions served, clicks, conversions, revenue, audience behavior, audience profile data, etc. At MediaMath, we are always looking to enhance our cutting edge infrastructure. We were looking to take our existing capabilities to the next level to manage new innovative analytics tasks. Processing this raw data to segment the audience, optimize campaign yield, compute revenue attribution, etc., is a non-trivial problem for some of the following reasons:

1. Complexity of transforming Semi-Structured data

Transforming session log data to construct user sessions and click-path analysis for further analysis is a complex process. We knew that Apache Hadoop was an attractive alternative but we wanted a solution that our analysts could easily use and get started with quickly and did not have to worry about the operational management of such technical options. We wanted a solution where analysts could focus on their data and transformations without having to think about issues such as cluster sizes, Apache Hadoop versions, machine types and other elements of cluster operations.

2. Data Pipelines

We needed a service to develop data pipelines that repeated the same transformations, day-after-day, week-after-week, without much intervention from my team, once it was setup. Automating the execution of the data pipeline, while honoring the interdependencies between the pipeline activities was a crucial requirement! We had learnt our lessons via prior experiments with cron that this wasn't the best approach.

3. Low risk Apache Hadoop

We needed something that was reliable and easy to learn, setup, use and put into production without the risk and high expectations that comes with committing millions of dollars in upfront investment.

The Solution

We evaluated a few Apache Hadoop based offerings and decided to give Qubole a try:

1. Big Data Analytics Solution - During our trial, we quickly created an account on Qubole and the team helped us upload sample data. We started using the system and immediately started to see the value of it. Within hours, we were able to re-use a number of very useful, business-critical, custom Python libraries that we had developed, matured, and stabilized. These libraries computed revenue attribution by customer and by campaign by mashing together semi-structured and relational data, as well as other useful tricks.

2. Cloud - We also noticed that the cloud-based Qubole clusters automatically grew the number of compute nodes as we started to run more queries and scaled the cluster down as the number of queries went down. This operational efficiency was a plus as we didn't have to continually reach out to our partners in Engineering who have the complex task of managing our mission critical production systems.

3. Data Pipelines - Qubole’s engineering team worked with our team to build a custom data collector from our Oracle Database to my Amazon S3 account. Using their S3 Loader and Sqoop-as-a-Service offering, they setup a pipeline that loaded the S3 data into Qubole’s Big Data Analytics Solution, did all kinds of processing, and pushed the resulting summaries into a MySQL instance that both our customers and we could query using our BI tools. We were set up and running in a few days.

4. Risk Free - Qubole‘s interfaces, including its easy to use GUI that really simplifies big data and its support for SQL with easy ways of embedding custom libraries, made it easy to learn. Using their GUI, setting up and tearing down clusters was totally transparent -- as an analyst I did not have to take on such an operations headache. We saved the company a few million dollars of upfront investment by going with Qubole. Also, the Qubole guys are a seasoned bunch who seem know what they are doing, and have credible answers and solutions to the team’s questions. They are a Skype-chat or a phone call away whenever my team needs help with issues or change requests. I don’t feel I am taking on a huge risk by going with Qubole. Over time, they have become a partner in my team’s success, one to whom I delegate my big data platform needs.

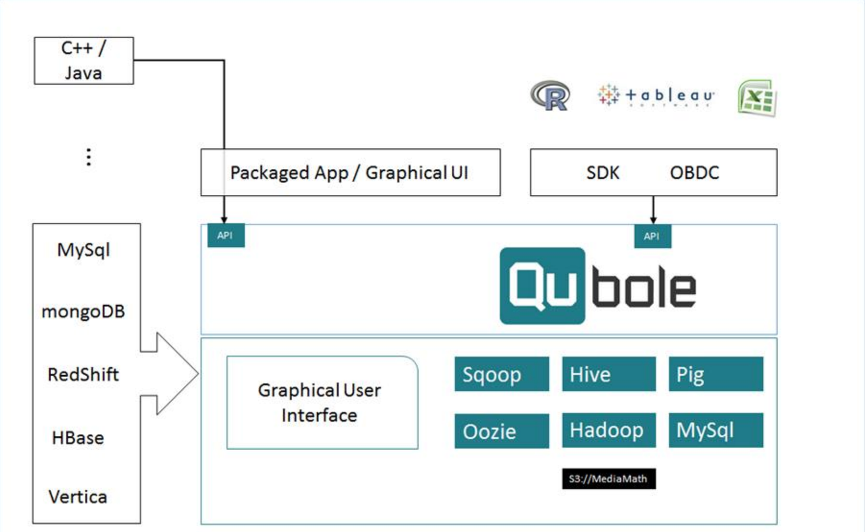

Figure 1. Sample Architecture

“I am very happy with Qubole! Our goal at MediaMath was to take our existing industry leading infrastructure to the next level handling new complex analytics tasks. Qubole has helped us enable this goal with minimal risk.”

About Qubole

Qubole is the leader in Big Data SaaS. Qubole was founded by Ashish Thusoo and Joydeep Sen Sarma, former leaders of Facebook's data infrastructure organization and long-time contributors to Apache Hadoop and creators of Apache Hive. Qubole is trusted by the largest brands in social media, online advertising, entertainment, gaming and other data-intensive ventures.

Qubole Data Service (QDS) provides a turnkey SaaS solution for Big Data teams, running on the top performing elastic Hadoop engine on the cloud and includes a library of data connectors with graphical user-interface for Hive, Pig, Ooze and Sqoop. QDS makes it easy to inspect data, author and execute queries, and convert queries into scheduled jobs. With QDS, the power of Big Data meets the simplicity of the cloud.

The World's Biggest Brands Run on Qubole:

saavn TubeMogul Pinterest

Nextdoor PUBSQUARED MERKLE A Database Marketing Agency

Quora MediaMath

To try Qubole for free for 15-days, go to http://www.qubole.com or email sales@qubole.com

855-HADOOP-HELP