



Return Path Accelerates Data Science Projects, Saves Thousands In Compute Costs With Qubole

Exploring the Business Problem

The data science team at Return Path is chartered with maximizing the economic value of the company's data. To accomplish that objective, the team established two high-level deliverables:

1. Generate data-driven insights that improve business decisions.
2. Ensure customers' problems and support questions are addressed in a timely manner.

However, the company's complicated on-premises technology stack included diverse tools and multiple data silos and repositories that made servicing data requests difficult. The process was so arduous that data scientists were spending up to **80% of their time** on internal and external data requests. With the complex infrastructure and overhead of accessing siloed information, data scientists found it challenging to deeply explore data, uncover hidden insights, and pinpoint new business opportunities — which ultimately created bottlenecks and extended the time to market for data science projects.

A Vision for a Unified Infrastructure

Executives at Return Path issued a goal to consolidate their infrastructure and technology stack on the cloud, which had become an intricate web of tools and applications spanning multiple acquisitions and teams. The company selected Amazon Web Services as their cloud service provider, in part because a portion of infrastructure was already hosted on AWS as the result of a previous company acquisition. In conjunction with infrastructure alignment, the company also sought a means to deliver self-service access to internal and external data consumers.

About Return Path

Return Path is an email intelligence company with over 18 years in the industry. By leveraging its data partnerships and consumer panels, the company helps its customers optimize email marketing programs to reach the right people with the right message at the right time.

“Qubole helped prevent us from making bad decisions that cost the business tens or hundreds of thousands of dollars.”

Robert Barclay

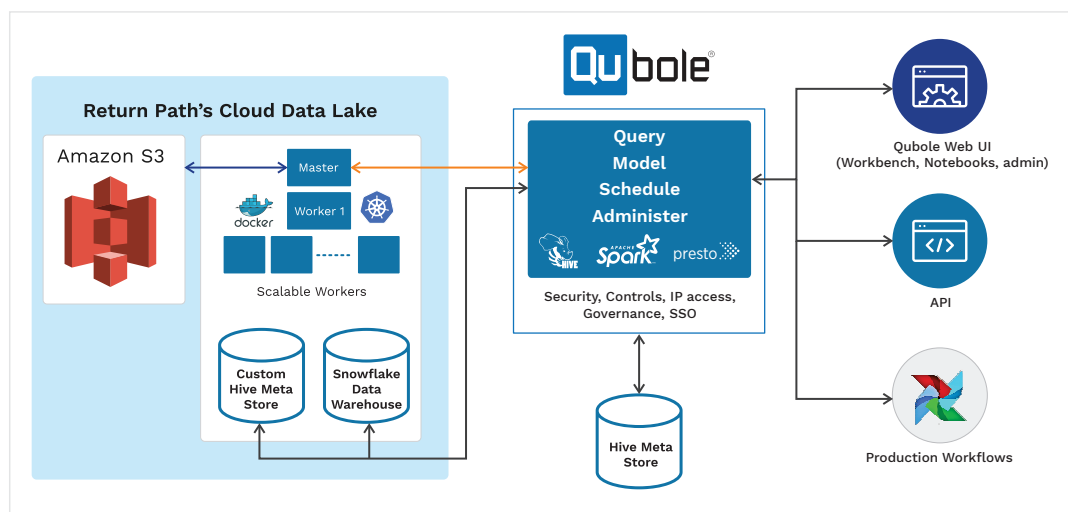
VP of Data and Analytics, Return Path

Driving Down Compute Costs with Qubole

As the data science team began moving data workloads to the cloud, they saw their AWS compute costs balloon. Team leaders sought a tool to help control expenses, and found value in Qubole's out-of-the-box big data cloud infrastructure management capabilities such as autoscaling and spot buying. In addition, the platform's collaborative Notebooks and Analyze features met the team's requirements for enabling self-serve access to data users and ensuring that the data science team was no longer the bottleneck.

Enabling Self-Service Data Access

One pressing initiative for the data science team was enabling data access for all relevant users. The team's previous strategy was to create reports for internal teams as well as customers to provide insights for business decisions. Yet report building often consumed a considerable portion of the team's time, because single reports did not always answer the question being asked or provide the depth of detail needed. With Qubole, internal data users can now directly explore and query the data to generate hidden insights and uncover new business opportunities while reducing support overhead on the data science team — allowing them to focus on more strategic projects leveraging machine learning and predictive analytics techniques.



Qubole's support for multiple engines, analytics, and dashboards, as well as the ability to access multiple databases, allows Return Path to create a unified view of their data and collaborate across multiple user personas — technical and non-technical.

Tackling Machine Learning Algorithms

Email message classification was another data science project that Return Path tackled more easily with Qubole, since it allows Return Path and its customers to understand at a granular level the success of email marketing programs. The project emerged as part of a business goal to apply machine learning (ML) to reduce the time spent generating partner reports on campaign types. Previously, the services team spent 50-plus hours per customer manually tagging and aggregating data for this report. The figure below shows the raw data in contrast with the final report.

“With Qubole, I was able to automate over half of the volume that was coming into our team's support queue, and that freed up a lot of time for us to focus on more interesting problems.”

Sasha Mushovic

Data Scientist, Return Path

date	inbox	spam	email_read	trash	moved	tis	tins	delete_unopened
2018-01-01	0.994	0.006	0.445	0.143	0.14	0.0	0.0	0.052
2018-01-02	0.994	0.006	0.42	0.177	0.14	0.0	0.0	0.071
2018-01-03	0.864	0.136	0.152	0.163	0.083	3.7E-4	6.0E-4	0.115
2018-01-04	0.991	0.009	0.45	0.197	0.114	8.1E-4	0.0	0.099
2018-01-05	0.996	0.004	0.421	0.165	0.136	0.001	0.0	0.079

Classification	inbox	email_read	spam	trash	moved	tins	tis	delete_unopened
Birthday	0.942	0.24	0.058	0.231	0.098	0.02423	0.00255	0.119
Abandonment	0.946	0.205	0.054	0.236	0.083	0.00926	0.00105	0.128
Loyalty	0.936	0.215	0.064	0.236	0.089	0.02467	0.00172	0.124
Newsletter	0.903	0.182	0.097	0.212	0.108	0.01773	0.00305	0.117
Account	0.925	0.299	0.075	0.2	0.115	0.0132	0.00224	0.097

Leveraging the new cloud infrastructure, data scientists set up an automated workflow in Airflow to aggregate and consolidate employee-tagged data. When initiated, the process uses a boosted K-means algorithm to cluster the data by subject similarity to facilitate manual tagging. Employees receive an email with the data, then tag each entry by campaign type and upload the data into the repository. The data consolidated into a tagged data set was then used to build the classification model. Qubole Notebooks enabled the team to easily create the model and prototype different ML algorithms such as neural nets, regression, and decision trees — while using hyperparameter optimization techniques such as Grid Search, Hyperopt, and Random Search — then select the option best suited for each project.

The interactive dashboard on Qubole Notebooks allows for cross-functional collaboration and enables data consumers in the organization to access the same environment — under different user roles — to test and provide feedback. Insights generated from this additional level of exploration help the data scientists at Return Path to better understand the industries and customers they serve.

Uncovering Future Data Science Projects

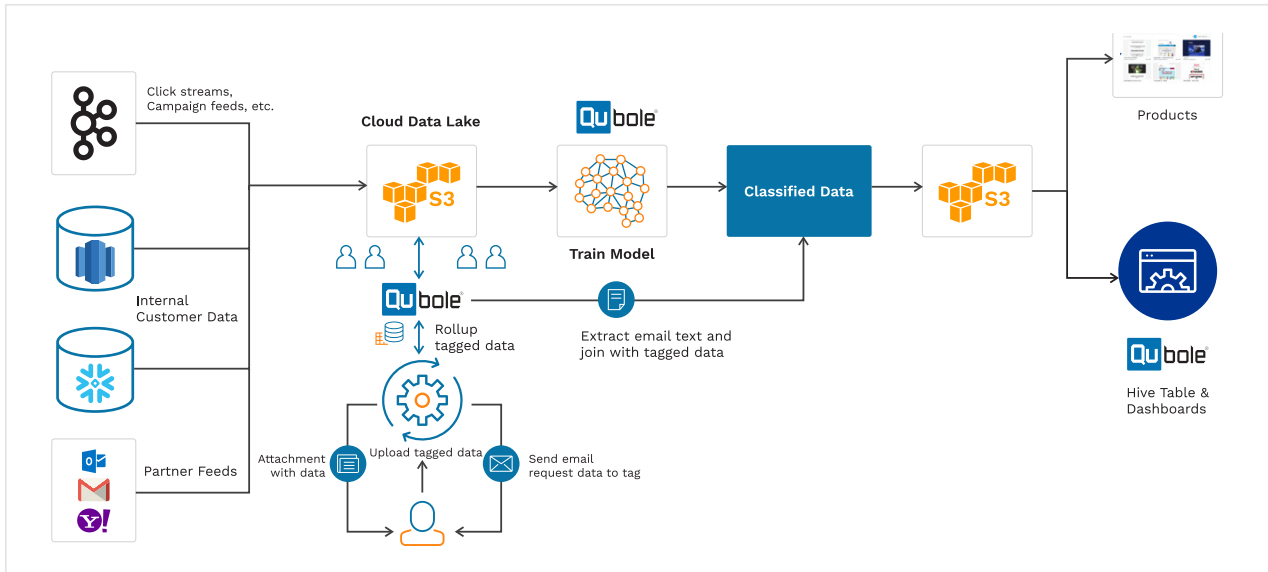
As the data scientists at Return Path explore their now fully accessible data set, they continue to discover new and valuable use cases that help the company improve or enhance existing services, as well as innovate to deliver new value-added offerings. For example, the team was able to develop a new product that allows Return Path’s customers to perform advanced email filtering based on many attributes, such as campaign type and draft marketing messages with the highest chances of success.

“ The way Qubole manages the infrastructure allows me to develop in other areas of the data product life cycle. I can spend more time on discovery, talking with clients, and really understanding their problems at a deeper level. ”

Sasha Mushovic

Data Scientist, Return Path

With a common platform like Qubole to collaborate on, Return Path’s data scientists, data analysts, and product managers are able to explore their data more deeply and uncover greater business value.



The Business Value of Qubole at Return Path



Reduced costs

- Workload-aware autoscaling and Intelligent spot management
- Self-service analytics for all data users and executives



Greater innovation

- Deeply explore data to uncover new business opportunities



Improved productivity

- Reduced time to deployment and 50% reduction in data preparation tasks



Increased customer satisfaction

- Fewer missed SLAs and faster turnaround on data requests

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. Qubole investors include CRV, Lightspeed Venture Partners, Norwest Venture Partners and IVP. For more information visit www.qubole.com

FOR MORE INFORMATION

Contact:
sales@qubole.com

Try QDS for Free:
<https://www.qubole.com/products/pricing/>

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

WWW.QUBOLE.COM