Industry: Media

TrafficGuard Halts Digital Ad Fraud with Qubole

What Is Ad Fraud?

Ad fraud relates to generating invalid traffic —such as impressions (ad views), clicks, or app installs— to either fake legitimate advertising engagement, or steal attribution of legitimate advertising engagement from other sources.

Ad fraud costs advertisers—who typically pay by the impression, click, or install event—billions of dollars per year. Indeed, according to Juniper Research, the global cost of ad fraud will top $44 billion by 2022. (As far as criminal enterprises go, only the international drug trade is more lucrative.) Ad fraud also reduces the effectiveness of advertising campaigns and erodes trust among advertising clients.

Advertisers have traditionally addressed ad fraud through a reactive process of reconciliation at the end of a set period, typically a month. This involves identifying ad fraud long after it has occurred and seeking reimbursement for any associated ad spend. But this process is costly, time-consuming, and error prone—and does nothing to deter fraudsters or thwart future fraud.

About TrafficGuard

TrafficGuard is a software as a service (SaaS) product that detects, mitigates, and reports ad fraud in real time through the use of several sophisticated machine learning (ML) models. TrafficGuard is owned by Australian-based Adveritas Limited, which specializes in big data technologies and Ad Tech software solutions.

Find more information about TrafficGuard at https://trafficguard.ai

A better approach is to prevent ad fraud from occurring in the first place and stemming the flow of ad spend to fraudsters. That's what TrafficGuard does. It discourages fraudsters by detecting and preventing ad fraud at several stages in the advertising journey including the impression (when ads are viewed), the click, and as events occur, such as app installs. "The idea behind TrafficGuard is to utilize data at every stage of the advertising journey to block advertising fraud as early as it can be reliably detected," says Head of Data Science and Analytics, Raigon Jolly.

Business Problem Overview

"Ad fraud is trying increasingly to resemble human behavior," says Jolly. So, "in order to reliably detect something that is continuously evolving, the technology must be very sophisticated. We have a rapid rate of development so we needed tools that simplify and streamline infrastructure management so that we can focus on developments that our clients directly benefit from.

TrafficGuard relies on big data processing to detect and prevent ad fraud, which requires a robust infrastructure. So, early in the product-development process, TrafficGuard turned to Qubole to provide a cloud-native data processing infrastructure that would guarantee the scalability and cost efficiencies required.

Assembling the Infrastructure

Detecting and preventing ad fraud in near-real time means processing considerable amounts of data. Rather than attempting to build and manage the architecture themselves, the TrafficGuard team opted to partner with Qubole from the start.
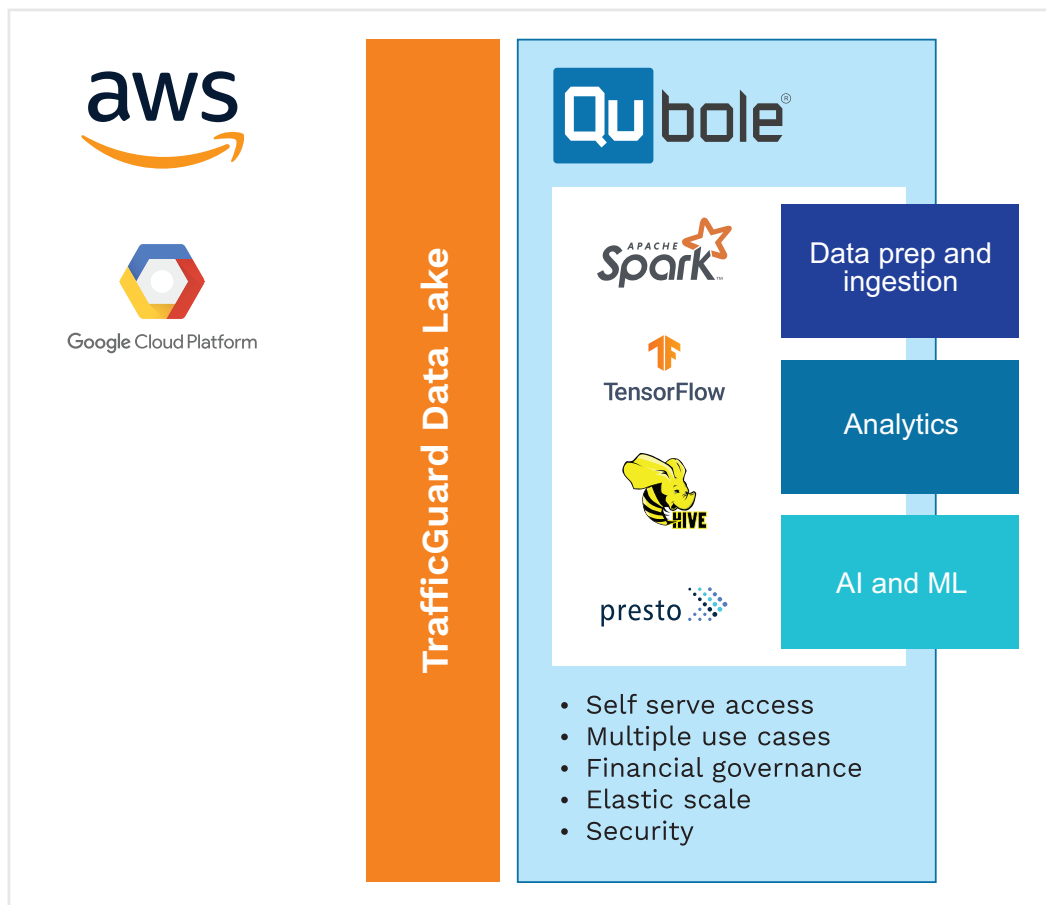
Senior engineer Pravin Todkar says that Qubole enabled TrafficGuard to bring their innovative fraud prevention to market much more quickly. "It would have been really difficult to build TrafficGuard without Qubole," he observes. Moreover, Qubole empowered the TrafficGuard team to "focus on product innovation rather than infrastructure management," says Todkar. This has yielded concrete benefits, like quicker development turnaround times. It has also resulted in less tangible rewards, such as fostering a culture of innovation—which in turn inspires team members to take risks and helps attract and retain top talent. Faster time to market and an innovative culture—that's a recipe for success.

Boosting Efficiency and Keeping Costs Down

TrafficGuard processes approximately 1 billion data transactions a day—roughly 10 terabytes and is rapidly scaling, over the last 6 months they have seen a 12 X increase. Some of these data transactions are essentially constant and require always-on clusters. But other data transactions spike at unpredictable times. For these, TrafficGuard employs AWS Spot instances through Qubole's Intelligent Spot Management capabilities. "Spot instances are really helpful," says Todkar. "With spot instances, it takes only a second to spin up clusters and start running workloads."

Spot instances improve efficiency and keep costs down. They achieve this by aggressively downscaling as soon as the workload is complete—meaning that TrafficGuard never pays for idle clusters. And because all this upscaling and downscaling occurs automatically in Qubole—based on workloads, job priority, or SLAs—the company saves on labor costs, too. That is, rather than taking on new hires to manage all these (and other) operations, "we've been able to achieve more with the DevOps resources we currently have," says Jolly.

On top of all that, says Todkar, "Qubole has an excellent customer support team. Their expert opinion comes in handy at times when we face technical issues."

> *Qubole has helped us federate data, manage data pipelines, streamline infrastructure management and leverage open source technologies to support our efforts in building enterprise-grade ad fraud prevention.*

Pravin Todkar
Senior Engineer, AI, Data, and Cloud Solutions

> *We are a growing business with new clients coming onboard frequently. The nature of digital advertising is that traffic volumes can be volatile, fluctuating with little warning. For fraud detection and other business needs, Qubole handles fluctuations of data with autoscaling. Spot instances are really helpful in terms of managing cost.*

Pravin Todkar
Senior Engineer, AI, Data, and Cloud Solutions

Working with Other Tools

Qubole serves as the foundation for TrafficGuard, powering various technologies such as Apache Spark (for processing data jobs), Apache Airflow (for managing data pipelines), Presto and Hive (for analytics), among others. But TrafficGuard requires several other tools, either integrated with or downstream from Qubole—like Druid, Elasticsearch, Redis, Tableau, and other analytics and AI frameworks from Google Cloud Platform (GCP) and Amazon Web Services (AWS)—to deploy machine learning models that detect ad fraud, generate reports for the TrafficGuard team and its clients, and deliver proactive fraud detection alerts.

Looking Ahead

Many solutions designed to combat ad fraud rely on tools like rules engines and IP blacklists to detect it. But, these tools present two critical limitations. First, these tools look for known indicators of ad fraud, so are not suitable for detecting new fraud tactics as they evolve. This leaves advertisers exposed to new forms of fraud. Second, these tools may flag valid impressions, clicks, or install events as fraudulent. These false positives can result in valid traffic being removed, legitimate supply sources not receiving due payment, and advertising campaigns' effectiveness compromised.

To overcome these limitations, the TrafficGuard team has turned to sophisticated machine learning models. Management of data pipelines and infrastructure to support this effort is streamlined with Qubole. These models analyze combinations of indicators over time and across devices to detect fraud as it evolves as well as mitigate false positives. When used in concert with rules engines and blacklists, TrafficGuard's models provide far greater protection against both known and unknown forms of ad fraud.

So far, the TrafficGuard team has developed more than 10 machine learning models for use in fraud detection—a number "that is expected to increase significantly in the near future," says Todkar.

> *Given our fast pace of growth, our data needs are ever increasing. Scale and reliability are integral to our solution, as is continuous innovation to keep ahead of sophisticated fraud. Using Qubole, our infrastructure automatically adapts so the team can focus on developments that our customers can see.*

Pravin Todkar
Senior Engineer, AI, Data, and
Cloud Solutions

Business Value

- The ability to bring TrafficGuard to market much more quickly than by using legacy architectures

- The ability to focus on product innovation rather than infrastructure management—fostering a culture of innovation, and helping to attract and retain top talent

- The ability to process 1 billion data transactions (roughly 10 terabytes) each day and keep costs low through Qubole's workload-aware autoscaling, heterogeneous cluster and intelligent Spot management capabilities

- Time savings and trust developed thanks to Qubole's excellent customer support

- The ability to integrate and leverage multiple open source engines, frameworks, and third-party tools to power TrafficGuard's solutions

- The ability to prepare data and to train and deploy the multiple sophisticated machine learning models required by TrafficGuard detection processes

- Managing multi cloud data pipeline

About Qubole

Qubole is revolutionizing the way companies activate their data — the process of putting data into active use across their organizations. With Qubole's cloud-native big data platform, companies activate petabytes of data exponentially faster, for everyone and any use case, while continuously lowering costs. Qubole overcomes the challenges of expanding users, use cases, and variety and volume of data while constrained by limited budgets and a global shortage of big data skills. Qubole offers the only platform that delivers freedom of choice, eliminating legacy lock in — use any engine, any tool, and any cloud to match your company's needs. Qubole investors include CRV, Harmony Partners, IVP, Lightspeed Venture Partners, Norwest Venture Partners, and Singtel Innov8

For more information visit **www.qubole.com**.

FOR MORE INFORMATION

Contact:
sales@qubole.com

Try Qubole 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