CASE STUDY

Industry: Technology

Predicting, Detecting, and Eliminating Online Threats: Malwarebytes

Business Problem Overview

To predict, detect, and neutralize emerging threats, Malwarebytes processes billions of threat telemetry records daily. The company then performs advanced analytics on this data to identify potential threats and runs ML and AI models to determine what action to take to protect its customers.

Malwarebytes formerly relied on a third party on-premises deployment to ingest and process this data. But this system proved inadequate. For example, the pipeline took a few days to complete Extract-Transform-Load (ETL) on one data stream alone. And, queries on the ingested data were painfully slow.

That wasn't all. It was also expensive—and was becoming increasingly more so as Malwarebytes' data grew exponentially. At the same time, little was offered in the way of support. "We started getting into issues where we were all on our own," says Malwarebytes' Senior Manager, Data Engineering - Data & AI, Sujay Kulkarni.

Malwarebytes needed some way to modernize its big data processing to improve turnaround time while also keeping costs down. And the company needed more than just a vendor to support this operation—it needed a partner. So, in 2016, it turned to Qubole.

About Malwarebytes

Malwarebytes is a cybersecurity company that produces anti-malware software for a variety of platforms. The company offers consumers free, premium, and enterprise-grade versions of Malwarebytes, which detect, remove, and remediate computer malware. Malwarebytes uses machine learning (ML) and artificial intelligence (AI) to identify and predict emerging threats before they infect machines.

Qubole's Elasticity Improves Processing Speed and Lowers Costs

Today, the scale of Malwarebytes' threat data processing and analysis is massive. The company focuses on helping its consumers and business customers to protect against threats, not just remediate malware infections. It uses this data to identify anomalies based on endpoint variations and critical detections. But this data is not only used for predicting potential threats.

> *"We really needed to work with a company that had mastered the elasticity of the cloud, and that company is Qubole."*
>
> **Manju Vasishta**
> Director of Data Science and Engineering
> Malwarebytes

Malwarebytes adopted Qubole in concert with Kafka (for ingesting data streams), and an AWS S3 data lake (for data storage). The virtually unlimited compute power in the cloud seemed like the solution to its performance problems, but soon proved cost-prohibitive. Compared to on-premises, adding compute was fast, as administrators could do it themselves at any point. But manually releasing it when it was not needed was almost impossible due to the bursty and unpredictable nature of big data workloads. For example, a ransomware outbreak in the world, resulting in a surge in data volumes of 5x to 25x or more. This resulted in constant cost overruns and fire drills for overloaded systems administrators.

Malwarebytes' data processing paradigm changed with Qubole. First, it de-coupled compute and storage. Second, "playing by the rules of the game of the cloud," says Kulkarni— leveraging things like autoscaling (scaling out and scaling up, and being elastic & ephemeral in nature), low-cost compute instances (AWS Spot), and storage (an AWS S3 data lake)— significantly improved the efficiency of data platform. Today, Malwarebytes uses Qubole to process its data. About 60 to 70% of it is logs, telemetry and other types of unstructured and semi-structured data that is being processed in Qubole.

"Qubole has really mastered the elasticity component of the cloud," says Malwarebytes Director of Data Science and Engineering Manju Vasishta. "Qubole helped us run our ETL at night, spinning up and spinning down clusters when we needed them." This ability to add and remove compute resources on demand—based on the workload or SLA, and without human intervention—in a matter of minutes has greatly increased the speed at which Malwarebytes processes critical data, directly affecting the company's ability to detect, predict, and remediate emerging threats.

Qubole isn't just quick. It's highly efficient, too. "You really have to aggregate the heck out of our data to make any sense of it or to bring it to a level where it can guide us in our decision-making process," says Sujay. He cites one key project for which Qubole aggregates and processes between 20 and 48 terabytes of raw data per day but delivers just 2 to 3 terabytes of meaningful and actionable data. Qubole provides a single framework for processing data more quickly, whether for use in ML models for predictions, in BI applications for business reporting, or for GDPR compliance—all with just one full-time administrator plus three senior engineers—a few times per quarter. The result is more powerful insights, because they involve better data.

Finally, Qubole is cost-effective. Malwarebytes pays Qubole only for the resources it uses.

Easy Adoption and a Quick ROI

For Malwarebytes, adopting Qubole proved painless. "It was easy to jump on board," says Kulkarni. "The technology was fantastic. We got the appropriate help. Native integrations with modern cloud services out of the box got us going. The early part of the journey was frictionless. We had a fantastic experience with the folks on the Qubole side." And, he adds, Qubole features like intelligent **Spot bidding** were "something we could quickly implement and start showing an ROI very soon, without having to completely change the way we did things."

The **platform's ROI** was quickly revealed in another critical way: by the meaningful data it helps to discover, which yields more powerful insights. These insights—for example, predictive insights about emerging threats; marketing lead conversion propensity using ML algorithms; behavioral clustering of malware; sentiment analysis of reviews about Malwarebytes products and features on various social media platforms using advanced natural language libraries; have driven key decisions that have served the business and its customers well.

Ease of Administration and Top-Notch Support

Before Malwarebytes adopted Qubole, it often found itself buried with administrative tasks. "The rate at which we were growing meant that we had to gradually but constantly provision more human and computing resources. We had to buy more licenses, and always redistribute and rebalance the cluster," says Kulkarni. "It was not scalable, not economical, and it ate a lot of man-hours." But with Qubole, he says, "this type of administration takes up about 1% of our time because it's mostly automated."

> *"We believe in having partners, not vendors. That's what Qubole is: A Partner."*
>
> **Sujay Kulkarni**
> Senior Manager, Data Engineering - Data & AI, Malwarebytes

Qubole also helps facilitate another aspect of administration: **GDPR compliance**. "Qubole was instrumental in our efforts to comply with GDPR. It saved us time, effort, and money, and gave us the peace of mind that our data processes were compliant," says Kulkarni.

Perhaps most importantly, Qubole continues to provide top-notch support—With Qubole, there are people who specialize in the tools we use. So, they don't just help us resolve current issues, they help us to focus on the future."

Looking Ahead

Malwarebytes' initial needs required Qubole primarily for building data pipelines for downstream uses, "but we have started using Qubole for more data science–focused activities," he says—things like modeling and analytics. And, he says, "Qubole's notebooks are very good." Vasishta is also intrigued by Qubole's serverless query engine, called Quantum. "I'm interested to see if we can use Quantum to do serverless SQL querying, particularly for occasional users," he says.

> *"Data science and SQL as a service—those are two things we are actively looking at using Qubole for."*
>
> **Manju Vasishta**
> Director of Data Science & Engineering
> Malwarebytes

Business Value

- Workload and SLA-aware autoscaling of compute resources yields greater data-processing capacity at much lower costs
- Improved efficiency produces meaningful data and more powerful insights
- Easy user on-boarding resulting in high adoption
- Quick, tangible ROI
- Ease of administration aided by top-notch support

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