Qubole® / Fanatics

How e-tailer Fanatics cut cloud costs by 70% and development time by 60% with Qubole

In supporting the websites of five sports merchandise brands of its own and more than 600 partner sites globally, Fanatics' cloud-based e-commerce platform handles billions of data points daily.

Raj Tanneru, Director of Data Engineering at Fanatics, estimates they process an average of five terabytes of data every day, seven days a week. To maintain Fanatics' competitive advantage, Tanneru and his team are constantly building and modifying data pipelines to better support:

- Visitor-specific product rankings
- Personalized email product recommendations
- Site feature experimentation
- Ad hoc queries
- Business reporting

Lowering the cost of data processing is a high priority at Fanatics. That's why they turned to Qubole. Qubole is the open data lake company that provides an open, simple and secure data lake platform for machine learning, streaming analytics, data exploration, and ad-hoc analytics.

About Fanatics

Fanatics is the global leader in licensed sports merchandise. They offer the largest collection of timeless and timely fan merchandise, whether shopping online, on your phone, in stores, in stadiums, or on-site at the world's biggest sporting events.

Fanatics' innovative, tech-infused, multi-channel approach to making and selling sports merchandise in today's on-demand culture is transforming the way fans purchase their favorite team apparel and jerseys. Their brands – Fanatics, FansEdge, Kitbag, Majestic and Fanatics Authentic – are uniquely positioned to serve the ever-growing real-time appetite of fans worldwide.

Fanatics' goal is to be the premier global cloud commerce platform, supporting sports merchandise e-commerce sites worldwide.

Personalizing product relevance to each individual customer

One of Fanatics' biggest data tasks is sorting and ranking products from hundreds of catalogs to present each site visitor with the items that are most relevant to them.

Relevance and ranking can be based upon the individual's purchase history, search history, customer profile (location, favorite teams, etc.) and product popularity, as well as the visitor's search terms. Using Qubole, Tanneru's team can quickly build and modify data pipelines that integrate product scoring with machine learning models to provide each online shopper with an at-a-glance selection that's highly relevant to their desires, preferences and buying patterns.

Before adopting Qubole, Fanatics had a batch system in place for product scoring, but they were looking for ways to improve its efficiency. They experimented with a real-time scoring system but never took it to production. They felt an offline batch processing system worked better for them.

Switching to Qubole not only saved Fanatics time in building their scoring pipelines. They also gained greater insight into their data and greater cost effectiveness from their computing.

Greater efficiency in experimentation

Another area in which Qubole helps Fanatics is in testing proposed new features for its sites.

Before promoting any new site feature, Fanatics does extensive A/B testing[1] against a series of key performance indicators (KPIs). These KPIs include metrics like conversion rate, margin, and revenue per visit. Prior to adopting Qubole, Fanatics had to explore its test data with a third-party tool and build reports manually. That caused significant delays in decision making, aside from the interruptions in the interpretation and refinement of tests.

Now, Tanneru's team builds complex workflows within Qubole to generate the desired KPIs in a very short time. The experiments run by Fanatics' A/B testing team send site tracking information to their data lake in an hourly cycle. The pipelines, orchestrated in Qubole, automatically generate the KPIs in personalized reports, which help Fanatics' business team make near real-time decisions on the features. Tanneru says automation in Qubole has resulted in huge time savings in their experimentation.

> *We didn't have good success with AWS EMR in terms of cost effectiveness. That's one of the standouts of why Qubole works for us."*
>
> **Raj Tanneru**
> Director of Data Engineering, Fanatics

Generating email recommendations 60% faster at 25% lower cost

Fanatics also uses Qubole to send automated purchase recommendations by email to registered site users. Their recommendation engine not only factors in users' previous site experiences but also new products and other items the user may not have seen. It can even use real-time results from major sporting events—like the Super Bowl and other global sporting events—to capture sales from fan excitement and engagement.

Upon adopting Qubole, Fanatics moved their email processing and segmentation to Apache Spark on Qubole. This helped the company cut end-to-end processing time for this task by more than 60%—from thirteen hours down to five—and cloud costs by more than 25 percent.

Seeking greater efficiency

Because the data pipelines they construct are sophisticated, Tanneru and his team are always looking for new ways to increase efficiency. Tanneru says they're particularly interested in exploring RubiX, Qubole's Open Source data caching framework for cloud platforms. RubiX was developed to eliminate network I/O latency for frequently accessed data by caching it in local compute nodes. In the case of Apache Spark on Qubole, by reducing round trip calls to the object store, Rubix can improve its performance by as much as 300%.

Massive savings of both time and cloud costs

Reducing data processing time and keeping cloud costs under control are critical to the success of Fanatics' business. According to Tanneru, Qubole has been instrumental in both.

Qubole's simple, open, secure platform and near-zero administration—along with its APIs, pre-built integrations with 3rd-party

---

[1] 'A/B testing' is performing simple controlled experiments comparing two variants ('A' and 'B'). https://en.wikipedia.org/wiki/A/B_testing

solutions, and out-of-box tools for data science, data engineering and analytics—save Fanatics huge amounts of time in developing data pipelines, gaining insights into their data, creating business reports, and building machine learning models.

At the same time, thanks to Qubole's Adaptive Serverless Platform—which provides workload-aware autoscaling, automated cluster lifecycle management, and Intelligent Spot Management—Tanneru says Fanatics now uses about 80% AWS Spot instances against only 20% of On-Demand ones. By allowing them to use heterogeneous clusters with a higher percentage of lower-priced Spot instances in a highly efficient and automated way, Qubole has reduced Fanatics' overall cloud costs by more than 70%.

Solid support and game-changing solutions

Tanneru attributes much of the time and cost savings Fanatics have enjoyed to solutions suggested by Qubole's technical support team.

One such solution is Sparklens, Qubole's open-source tuning tool for Apache Spark. After being introduced to Sparklens, Tanneru says, he and his team were able to better focus their efforts on optimizing their data pipelines.

"Identifying problems like data skews isn't very straightforward using the native Spark UI. There's a DAG (directed acyclic graph), but in general, the visualization in the Spark UI isn't so helpful," says Tanneru. "Having a profiling tool—having real insights into the application—was something that was missing. With Sparklens, you can pinpoint problems and tackle them quickly. Qubole's Sparklens service has been a game changer for us."

> ❝ *Qubole's Sparklens service has been a game changer for us."*
>
> **Raj Tanneru**
> Director of Data Engineering, Fanatics

Tanneru also credited Qubole's support team with helping Fanatics save money by mitigating AWS Spot loss. He says Qubole introduced them to tools and techniques for identifying and optimizing AWS compute instances with higher AWS Spot loss. The recommendations also showed Fanatics how to fine tune Qubole's Intelligent Spot Management feature, which helps customers save up to 80% on cloud compute costs, regardless of the cloud of choice. "Qubole was really helpful and instrumental in some of the optimizations we have done here on our data pipelines," says Tanneru.