TiVo scales new media products with Qubole and data lakes on AWS

Unlocking data for targeted advertising and product improvement

TiVo's entertainment platform consolidates terabytes of data every month: raw viewership data from cable boxes in millions of homes, purchasing data from first and third parties, and location-based consumer data. TiVo's network and advertising partners need reports based on this data to better understand the viewing and purchasing behaviors of various customer demographics.

Because TiVo's partners often have drastically different reporting needs, all of this data needs to be transformed, segmented, and packaged in several different ways to satisfy their requirements. TiVo's data engineering team needed a way to do this efficiently, affordably, and at scale.

The need to process massive amounts of data efficiently and accurately

TiVo's existing approach involved ingesting data from several different sources into an ETL pipeline, which summed the data to Amazon S3. From Amazon S3, TiVo ingested data into various data marts and warehouses, ultimately processing it in Amazon Redshift and MySQL, and consumed the results using Java services on Amazon EC2. This approach required their data engineers to write new ETLs for each new report request, increasing dev time, inflating costs, and decreasing the overall efficiency with which reports were produced.

TiVo realized it needed a robust analytics platform that would allow it to scale and automate the process of ingesting, processing, and aggregating all of its disparate data while also driving down the cost of its analytics initiatives. To streamline its data science approach, TiVo would need a way to store all of its data – structured and unstructured – in order to remove data silos that prevented it from easily running the analytics workloads it required to generate requested reports.

QUBOLE'S BIG DATA ACTIVATION PLATFORM

- Constantly analyzes and learns about your data usage through a combination of heuristics and machine learning.

- Provides actionable Alerts, Insights, and Recommendations (AIR) to optimize reliability, performance, and costs.

- Automatically enacts policies through multiple Cloud Agents to eliminate the need for repetitive manual actions.

CHALLENGE
TiVo needed to make viewership, purchasing behavior, and location-based consumer data easily available for its network and advertising partners. It also needed the ability to quickly generate new report types for its partners without sacrificing developer hours.

SOLUTION
Qubole's cloud-native Big Data Activation Platform uses intelligent Cloud Agents to provide TiVo with critical data that its TV networks, programmers, and advertisers need to more effectively serve customers.

RESULTS
Qubole optimizes TiVo's analytics workloads, enabling the company to help its advertising and network partners reach their target customers through highly customized reporting.

The decision to use a data lake on AWS

To more readily make its data available for analytics operations, TiVo deployed a data lake on Amazon S3. The data lake allows the company to store any data type in a single convenient repository. Data can be collected from multiple sources and moved into the data lake in its original format. This allows TiVo to scale to data of any size, while saving time by eliminating the need to define data structures, schema, and transformations.

Why a Data Lake on AWS?

Flexibility and Agility	Security and Compliance	The Most Complete Platform for Big Data

aws

Activate your data on AWS, making it highly available for analytics

TiVo's data engineering team chose Presto as its query engine based on its flexibility and efficiency. The team then decided to use Qubole, which allows it to easily scale and manage its Presto clusters and more easily audit queries and debug commands. The Activation Platform provided out-of-the box functionality that Tivo would have needed to create from scratch if it had chosen to deploy Presto on top of AWS EC2 without Qubole. TiVo's data engineers found Qubole simple to deploy: after configuring permissions for AWS and the Qubole website, they were ready to run queries.

Qubole templates automate every element of TiVo's queries, including activating Presto clusters and scaling the clusters based on usage. This eliminates the need to manually write scripts to tell Presto how to behave. The query results are then saved in Amazon S3 buckets for later auditing. Through its service administration portal, TiVo can track its queries and view and download intermediate queries and results.

Qubole's rich feature set includes the ability to label individual clusters according to their workload. TiVo labels clusters (e.g. "ETL," "Reporting," and "Interactive") to help its team of developers stay organized. Qubole's notebook feature provides a convenient way to save, share, and re-run a set of queries on a data source – for example, to track changes in the underlying data over time, or to provide different views using different parameters.

" The Qubole interface makes it easy for our developers to go to a notebook, pick a cluster, and get started with a query. They don't have to worry about managing the cluster, and they're able to collaborate with other developers easily by sharing notebooks."

Lucas Waye, *Principal Engineer, TiVo*

Qubole®   **aws**

Qubole gives TiVo's partners the reporting they need

Qubole streamlines the process of generating reports for TiVo's partners, whose needs are constantly changing week-to-week in terms of scope, data type, and time (weekly, monthly, yearly). The financial and human resources required to run data science operations depend heavily on the complexity of the reports being run. Today, TiVo can do more with fewer resources by automating its reporting with Qubole.

Qubole provides a simple, intuitive way for TiVo's partners to set up and schedule reports tailored for their specific requirements. This self-service feature provides TiVo's network and advertising partners with the business intelligence tools they need to interpret data from highly targeted demographics at a cadence that works best for them. Having access to any kind of viewership, and purchasing only the reporting they need, allows networks and advertisers to more easily customize and scale new media products to thrive in a highly competitive space.
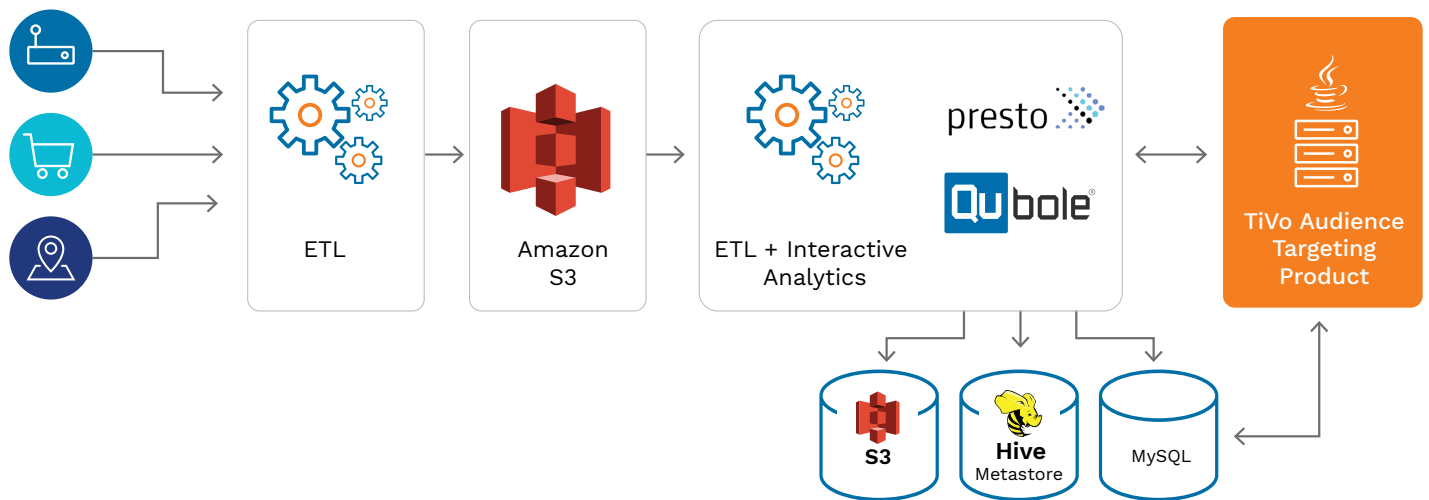
ABOUT QUBOLE

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 Activation Platform, companies exponentially activate petabytes of data 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's intelligent automation and self-service supercharge productivity, while workload-aware auto-scaling and real-time spot buying drive down compute costs dramatically. Qubole's platform delivers freedom of choice, eliminating legacy lock in—use any engine and any tool to match your company's needs.