

CLIENT Qubole
PROJECT Case Study: Zillow
OBJECTIVE Demonstrate the business value of Qubole's big data activation and processing platform for customers like Zillow

COPY EXCERPT

Powering the Complete Online Home Lifecycle with Qubole



Call or write CopyEngineer to receive a PDF of the complete case study.

Or view/download it online at:
www.qubole.com/resources/case-study/zillow

The Challenge of Eliminating Data Silos

Zillow's portfolio of brands is, quite literally, data driven. Their business model depends on putting up-to-date and accurate real estate information instantly at the fingertips of both consumers and real estate professionals. Keeping their businesses running smoothly – and improving them – depends on putting vast amounts of actionable information at the disposal of their business analysts.

As Zillow grew, the data volume of its various business units and the need to consolidate information across its brands has risen dramatically. In response, Zillow has been gradually migrating its business data from brand and application data silos to a group-wide data lake built on the Amazon AWS S3 platform.

But the Data Infrastructure team soon ran into several challenges processing data and managing its clusters due

to the technology it had originally selected. The inefficient compute autoscaling and Spot buying from the prior platform made it impossible to efficiently scale clusters for Zillow's highly variable Presto workloads and its huge clickstream batch jobs in Spark. Plus, frequent stability issues and lagging performance with Presto were putting a strain on the small Data Infrastructure team trying to meet its SLAs to Zillow's large user community.

Seeking Greater Performance, Stability and Economy

To address these challenges, Zillow turned to Qubole. The Qubole data platform supercharges productivity through intelligent automation, enabling a 1:100+ admin-to-user ratio, and delivers faster cycle times with a self-service infrastructure for all users. There is zero waste as Qubole's workload-aware autoscaling optimizes compute usage based on the task, SLA, or priority, thereby ensuring that a job will have the appropriate resources to complete. Both, aggressive downscaling and container packing also ensure there is no wasted dollars in compute when it becomes idle. Plus, the platform is always able to find the best (low cost) Spot instance buy to increase capacity temporarily while further driving down compute costs.

Shifting to Qubole also allowed Zillow greater flexibility. The platform is optimized for multiple clouds, including AWS, Microsoft Azure, Google, and Oracle, and it supports not only Presto and Spark, but also Hive, Airflow, Hadoop, Tensorflow and more. Plus, Qubole's customer-comes-first culture and industry-leading experience in both Big Data and the cloud meant they could get expert help in overcoming their cluster stability and performance problems.