We live in the age of big data, which allows organizations to turn data into insights in a timely manner. Organizations use it to drive innovation, inform targeted marketing campaigns, and become more efficient. Because big data requires combing, processing, migrating, and analyzing data from disparate sources and types, managing and running a big data platform is complex, time-consuming, and requires specialized personnel. At the same time, data keeps growing exponentially. Businesses demand a data infrastructure that is reliable, performant, and allows them to turn that data into insights in a timely manner. Organizations need a way to optimize their big data infrastructure, reducing time, cost, and complexity, to stay competitive.

A BIG DATA SOLUTION THAT’S OPTIMIZED FOR THE CLOUD

Qubole is a comprehensive, autonomous big data platform that self-manages, self-optimizes, and learns from your usage, allowing your data team to focus on business outcomes rather than on managing the platform. Qubole lowers the barriers to access, process, and analyze data, allowing data scientists and data analysts to get insights from their data faster. Qubole empowers you with its robust, cloud-native feature set that includes:

- Workload-aware auto-scaling that predictively manages cluster compute and storage size according to its workload.
- Automatic provisioning and termination of clusters - no need to manage infrastructure.
- The ability to intelligently shop and acquire Spot Instances to achieve an optimal combination of price, performance, and availability.
- Self-service data analysis that allows data teams to define policies that allow data access with minimal manual interventions.

A SELF-MANAGING AND SELF-OPTIMIZING DATA PLATFORM

Get more value from your data by enabling your data team to focus on business outcomes rather than managing big data infrastructure with Qubole, the autonomous data platform that:

- Constantly analyzes and learns about platform’s 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 a number of Cloud Agents™ to eliminate the need for repetitive manual actions.
A SINGLE PLATFORM FOR MOST USE CASES IN AWS DATA LAKE

Qubole is an ideal solution for processing, analyzing, and consuming your data in a data lake on AWS. A data lake allows you to conveniently store a wide range of heterogeneous data in a single cloud repository, where Qubole can automate and optimize your analytics workloads.

QUBOLE FOR ETL

Extract, Transform, and Load (ETL) tasks can be a major bottleneck for many companies. When ETL is generating roughly 80% of your compute hours, you’re constantly looking for ways to control and cut costs. Qubole’s data platform uses automated Cloud Agents to help data engineers focus on what matters most: helping create business outcomes instead of doing tedious, manual, repetitive tasks.

QUBOLE FOR AD HOC

Ad hoc analysis allows data analysts and data scientists to uncover the most critical insights for businesses. However, it’s not easy to size and configure technology to support ad hoc analysis, which can lead to unpredictable costs. Qubole auto-scaling allows ad hoc queries to run efficiently without incurring unpredictable costs. In addition, because Qubole offers a variety of ad hoc engines such as Hive, Presto, or Spark SQL, data analysts and data scientists can use data engines they are familiar with to deliver results more quickly.

A SELF-SERVICE PLATFORM FOR DATA SCIENTISTS

With Qubole, data scientists no longer need to rely on data administrators for provisioning compute clusters and resources. Once the exploratory data analysis and/or model development is done, data scientists can productize their notebooks using the notebook API with just a few clicks. Data scientists dramatically reduce the time to get machine learning into production. They use Qubole to experiment and determine the machine learning algorithm that better fits the problem. Then they prepare the data for the machine learning algorithm, train it, and put it in production, all without leaving Qubole.
POWERFUL ARCHITECTURE TO SIMPLIFY AND AUTOMATE EVEN THE LARGEST BIG DATA WORKLOADS

Qubole Data Service orchestrates resources on the customer’s AWS account based on data engineer-defined security policies and target workload SLAs. Separation of compute and storage ensures that compute resources can be scaled as needed without being constrained by storage costs and limitations. Clusters are automatically activated, scaled, and terminated in response to user queries submitted via Workbench, Notebooks, API requests from external applications, or third-party business intelligence tools, via ODBC/JDBC integration.

ALERTS, INSIGHTS, AND RECOMMENDATIONS (AIR)

Using a combination of heuristics and machine learning, AIR provides actionable alerts, insights, and recommendations to ensure:

- Workload continuity.
- High performance.
- Low reliance on cloud resources.
- Greater cost savings.

By automating lower-level, repetitive tasks, your engineering team can be less reactive to problems and more focused on directing better business outcomes. With Alerts, Insights and Recommendations (AIR), Qubole constantly analyzes metadata about infrastructure (cluster, nodes, CPU, memory, disk), platforms (data models and compute engines), and applications (SQL, reporting, ETL, machine learning) so you can better understand performance, usage patterns and cloud spend.

ABOUT AWS

For 11 years, Amazon Web Services has been the world’s most comprehensive and broadly adopted cloud platform. AWS offers over 100 fully featured services for compute, storage, databases, analytics, mobile, Internet of Things (IoT) and enterprise applications from 44 Availability Zones (AZs) across 16 geographic regions in the U.S., Australia, Brazil, China, Germany, Ireland, Japan, Korea, Singapore, and India. AWS services are trusted by more than a million active customers around the world – including the fastest growing startups, largest enterprises, and leading government agencies – to power their infrastructure, make them more agile, and lower costs. To learn more about AWS, visit http://aws.amazon.com.

Learn more about Qubole by visiting AWS Marketplace

QUBOLE – THE AUTONOMOUS DATA MANAGEMENT COMPANY

Qubole is passionate about making data-driven insights easily accessible to anyone. Qubole customers currently process nearly an exabyte of data every month, making Qubole one of the leading cloud-agnostic big-data-as-a-service provider. Customers have chosen Qubole because of its robust, innovative autonomous data platform. This cloud-based data platform self-manages, self-optimizes, and learns to improve automatically and, as a result, delivers unbeatable agility, flexibility, and TCO. Qubole customers focus on their data, not their data platform.