

Size of data and performance needed to scale

Quartile faced challenges in storing and processing data for multiple advertising channels due to crucial needs in reporting and consolidating sales data, including over 60 days of attribution. The previous architecture couldn't keep up with the size of data Quartile was processing. The team was batch processing over 10TB of data on a single job, which applied all transformations required for data reporting, causing server unavailability and late deliveries of data points. This process alone, which is responsible for improving the performance of ads, had severe performance problems, such as individual jobs running up to 7.5 hours every day.

Technology evolution, from legacy to modern data stack

As part of the evolution of the company, Quartile's data architecture has reached new levels of maturity, growing from traditional SQL databases running on Azure cloud to a new solution with the Databricks Data Intelligence Platform as the foundation. This modernization has had direct impact in several areas of the company and clear benefits for Quartile's customers: with Delta Lake there has been more data reliability, faster performance and reduced costs. When migrating the data from a traditional SQL database to Databricks, they saw a considerable reduction in data volume, mainly due to the optimizations of Delta with version control and Parquet compacting, resulting in storage reduction from 90TB to about 18TB.

The lakehouse architecture has enabled Quartile's data stack to evolve. This has been accomplished through several processes: by leveraging Databricks Auto Loader for incremental and efficient processing of new data files as they arrive in cloud storage; by using Delta Lake for its open format storage layer, which delivers reliability, security and performance on the data lake; by using Databricks SQL, which continues to bring data engineers and data analysts closer with easy-to-build queries and dashboards; and with Databricks Workflows, which connects all of the pieces together in a stable and scalable way.

To provide the best experience for customers, Quartile needs to be able to retrieve accurate data. This is an important challenge that is easier to handle with Spark User-Defined Functions, as the company uses the power of parallelism to break down processing into as many parts as needed. For their solution to scale, they utilize Terraform for deploying all workspaces, easily spinning up new clusters and jobs as well as ensuring the correct standards are being followed across the company.

Helping their partners run ads better

It is critical that Quartile's customers have a centralized solution in which they can analyze their sales, costs and other metrics related to their campaigns. At Quartile, they take advantage of the solid data engineering work and integrate Databricks with Power BI for embedding the dashboards directly in their portal. This helps provide a single place for clients to both configure marketing campaigns across several channels as well as follow up on performance changes while still maintaining a smaller data storage footprint that is 80% less expensive compared to the data stored in traditional data warehouses by leveraging Delta Lake's file format on object storage. The ability to combine data from all different channels has already helped several of their customers — SmartyPants, as an example, has grown over 100% since they partnered with Quartile.