



GANNETT

CASE STUDY

INDUSTRY: MEDIA

USA TODAY Publisher Gannett Puts Data at its Core to Open Doors for New Business Opportunities, Integrates Multiple Data Sources with Qubole

Core Business Problem and Future State

After more than 100 years of growth to 86 newspapers (and soon to grow to 110), by 2015 Gannett faced a common problem: what to do with new log-level data derived from digital operations, and how to consolidate multiple siloed data warehouses and “sources of truth” into a unified data lake? At the same time Gannett was well on its way to move its core computing infrastructure from on-site hosting to Amazon AWS.

The newly formed Gannett Data Platform team worked with an internal operations team to deploy a candidate architecture for a data lake on AWS, based on large always-running EC2 instances with on-device storage, but soon came to the realization that running a lake in the cloud as if it was an on-premises installation was not ideal; it was impossible to cost-effectively balance the needs of storage and compute resources: Gannett was either paying for resources not in use most of the time, or suffering the consequences of spikes in activity that exceeded the fixed capacity of the computing cluster. By the end of the year the team started to look for alternatives; specifically looking for the ability to separate storage from compute, and allowing the two to scale independently.

In early 2016 the Data Platform team discovered Qubole, and quickly identified it as the solution they were seeking. In late spring the team started the transition and by July - in less than 7 months from initial discovery - had completed the move to a flexible architecture with the “lake” residing entirely on AWS S3, while the computing “platform” being composed of EC2 instances that were scaled up and out - and down and in - as needed.

About Gannett

Founded in 1906, Gannett is a media and marketing solutions conglomerate, with annual revenues of over \$3 billion. Gannett is the US's biggest newspaper publisher and its flagship publication, USA Today, is America's top daily newspaper. The company produces over 300 other digital, mobile and print publications, including Newsquest titles in the UK, and provides digital marketing and advertising services to businesses across the US through its LOCALiQ division.

Single Cloud Platform Brings Rapid, Cost-Effective Data Growth

Qubole has enabled Gannett to rapidly scale up its use of data. In the first two years, Gannett grew its data volumes seven-fold, from 100 to 700 terabytes. Likewise, Gannett moved from hosting four computing clusters to around 25, and from having fewer than 10 data analysts to 40 users today. And it did this with just a two-fold increase in costs.

With no limits imposed by the cloud platform on the compute capacity, Qubole has given Gannett the freedom to take on any extra large-scale processing required, without impacting its normal operations. And Qubole's Intelligent Spot Management and Workload-Aware Autoscaling features have brought major benefits, enabling Gannett to cost-effectively manage its large data sets and 'bursty' data workloads. "Our number one reason for choosing Qubole was we wanted to take advantage of cloud economics: only pay for what you use," said Oskar Austegard, Gannett's Senior Director of Data Solutions. "Qubole's autoscaling and downscaling is definitely a huge cost saver, and the ability to isolate workloads to separate clusters is key to efficient operations."

“

Our number one reason for choosing Qubole was we wanted to take advantage of cloud economics: only pay for what you use. Qubole's autoscaling and downscaling is definitely a huge cost saver, and the ability to isolate workloads to separate clusters is key to efficient operations.”

Oskar Austegard

Senior Director of Data Solutions

Better Data Analysis Leads To New Products and Services

Qubole has enabled Gannett to not only gather more data, but analyze it much more precisely, down to individual user level interactions instead of using aggregated information. This is helped by Qubole's support for best-in-class big data products like Spark and Presto. To best use this new capability, Gannett has set up four teams of data scientists and analysts, whose improved understanding of advertising effectiveness and customer behavior has led to the company developing new services.

One key area is marketing optimization. Gannett has built data models that identify customers' lifetime value and predict, say, those who may be about to quit their subscription, to take preventive action. The company can now make better content recommendations and, in particular, analyze in detail which micro-segments of the population respond to adverts from its B2B customers.

Benefits From Faster Data Access and Standardization

With the advances of Qubole's Presto engine, Gannett is in the process of moving workloads from AWS Redshift to Presto accessing the lake directly. But the new end-to-end platform has already provided faster analytics, and corporate-wide data discovery and integration. Gannett has set up standardized reports across the organization and is moving to an environment of increased self-service reporting – “one approved way to get at that source of truth,” Austegard said.

Using Qubole and Presto, analysts can carry out ad-hoc data discovery much faster – querying an entire day's worth of data of around 70 million records across 300 dimensions in seconds. “That use case for analysts is quite powerful,” Austegard said, adding: “The ability to interactively extract answers from large and mostly unprocessed datasets are highlights of using Qubole and Presto.”

Looking Ahead

Via Qubole, Gannett plans to introduce additional data sources and expand its data science capabilities to continue improving its predictive understanding of user behavior and the content and advertising that appeals to them, in order to build new business lines and potentially open up new publishing and advertising markets.

Going forward, the goal is for data to have a place in every product decision: “How can we use our first party data and the insights derived from that data to improve the experience of both our B2C and B2B customers?” Austegard asks.

“

The ability to interactively extract answers from a large and mostly unprocessed datasets are highlights of using Qubole and Presto.”

Oskar Austegard

Senior Director of Data Solutions

Business Value



Rapid growth in data storage and utilization

- Data volumes up seven-fold in two years
- Extra processing performed without impacting normal operations



Reduce cost

- Intelligent Spot Management introduced
- Workload-Aware Autoscaling is a “huge” cost save



Better analytics feed new product development

- Data analyzed down to individual user-interaction level
- New marketing services launched through better understanding of customer behavior
- Better promotion of content and direct-sold advertisements



Faster data discovery and integration

- Analysts can query tens of millions of records in seconds
- Standardized reports introduced across organization
- Moving to an environment of self-service reporting

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 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 offers the only platform that delivers freedom of choice, eliminating legacy lock in — use any engine, any tool, and any cloud to match your company's needs. Qubole investors include CRV, Harmony Partners, IVP, Lightspeed Venture Partners, Norwest Venture Partners, and Singtel Innov8. For more information visit www.qubole.com.

FOR MORE INFORMATION

Contact:
sales@qubole.com

Try Qubole for Free:
<https://www.qubole.com/products/pricing/>

469 El Camino Real, Suite 205
Santa Clara, CA 95050
(855) 423-6674 | info@qubole.com

WWW.QUBOLE.COM