

acceldata

Customer Case Study

PubMatic Leverages Acceldata's Data Observability Platform to Optimize Performance and Cost at Massive Scale



Case Study: PubMatic

Problem:

- Excessive MTTR, frequent performance issues due to massive scale and large number of nodes in single cluster.
- High infrastructure and OEM support costs.

Solution:

Acceldata Pulse isolated bottlenecks, automated performance improvements, and distinguished between mandatory and unnecessary data to rapidly scale big data environment to meet expanding business requirements and reliably support mission-critical and customer-facing analytics requirements.

Results

- HDFS optimization reduced block footprint by 30%
- Kafka cluster consolidation saved infrastructure costs.
- Reduced OEM support costs to save millions of dollars/year in software licenses.
- Eliminated day-to-day engineering involvement and firefighting on outages and performance issues allowing data teams to stay focused on growing the business.



Recently public, fast-growing American ad-tech firm

Hyperscale setup with thousands of nodes handling hundreds of petabytes of data

HDFS, Yarn, Kafka, Spark, HBase

Open Source - HDP

~200 billion daily ad impressions, one-trillion advertiser bids /day, 2+ PB /day new data processed*

*As of December 2020

PubMatic is one of the United States largest AdTech companies. Since 2006, PubMatic has created an efficient global infrastructure with eight global data centers. The company is one of the industry's leaders in programmatic advertising innovation.

As of December 2020, PubMatic every day served 200 billion ad impressions, handled one trillion advertiser bids, and processed more than 2 Petabytes of new data.

Infrastructure/Environment

PubMatic is in hyper-scale mode. Its current environment includes 3,000+ nodes, 150+ Petabytes and 65+ open-source HDP (Horton Dataworks Platform) Clusters and is expanding rapidly. In addition, PubMatic uses other tools in the Hadoop big data stack, including Yarn, Kafka (50+ small Kafka clusters with 10-15+ nodes/cluster), Spark, and HBase.

Situation

Because of its massively scaled environment, PubMatic consistently experienced high MTTR (Mean Time to Resolution) metrics, frequent outages, and performance bottlenecks.

Many of the issues stemmed from its large numbers of nodes – in one case, 1,500 nodes in a single cluster. The system's instability resulted in time-consuming operational issues and constant daily firefighting. In addition, PubMatic was looking for ways to reduce its infrastructure and OEM support costs.

Business Impact

When PubMatic's data system performance wasn't able to keep pace with its rapidly-expanding business requirements, the company decided to implement a data observability platform to improve reliability, scalability, and the return on investment on its data operations.

The inability to correlate events across the infrastructure, data layers and pipelines meant that PubMatic could not materially improve its 'cost per ad impression' metric, which is one of its most critical performance metrics.

In addition, the company's rapid scaling resulted in unnecessary software licenses, which it felt could better align with actual needs. Finally, engineering's constant involvement in resolving operational system issues caused a distraction from the real objectives of scaling the data system to support the fast-growing business requirements.

Resolution

PubMatic began using Acceldata's Pulse product in mid-2020. At the data compute layer, Pulse immediately provided improved visibility into the inner workings of PubMatic's data applications and comprehensive observability for complex, interconnected data systems.

One of Pulse's most important benefits was its ability to predict, prevent and optimize PubMatic's data system performance at the very large scale that today's digital ad market requires.

In PubMatic's environment, Acceldata Pulse isolated bottlenecks and automated performance improvements. The product distinguished between mandatory and unnecessary data to ensure scaled growth that could reliably support all critical enterprise and customer-facing analytics requirements. Acceldata Pulse has helped PubMatic:

- ▶ Reduce 'cost per ad impression' - a key performance metric
- ▶ Improve reliability of data pipelines
- ▶ Eliminate day-to-day engineering involvement and firefighting by slashing the number of outages and performance degradation issues
- ▶ Decrease OEM support costs by \$10 million
- ▶ Optimize HDFS to reduce storage block footprint by 30%
- ▶ Consolidate its Kafka cluster and save infrastructure costs
- ▶ Saved millions of dollars in unnecessary software licenses



"Acceldata provided the data observability tools and expertise to make our data pipelines more reliable. They helped us optimize HDFS performance, consolidate Kafka clusters, and reduce cost per ad impression, which is one of our most critical performance metrics.

Acceldata's data observability saved us millions of dollars for software licenses that we no longer need. Now we can focus on scaling to meet the needs of rapidly growing business."

Ashwin Prakash

Engineering Leader, PubMatic