



CASE STUDY: PUBMATIC

PubMatic optimizes performance and cost at massive scale

| | | |
|---|---|--|
|  <p>Recently public, fast-growing American ad-tech firm</p> <p>Hyperscale setup with thousands of nodes handling hundreds of petabytes of data</p> <p>HDFS, Yarn, Kafka, Spark, HBase</p> <p>Open Source - HDP</p> <p>~200 billion daily ad impressions, one-trillion advertiser bids /day, 2+ PB /day new data processed*</p> | <p>Problems</p> <ul style="list-style-type: none">• Excessive MTTR, frequent performance issues due to massive scale and large number of nodes in single cluster.• High infrastructure and OEM support costs. <p>Solution</p> <p>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.</p> <p>Results</p> <ul style="list-style-type: none">▪ 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. | <p>“ <i>Acceldata provided the data observability tools and expertise to improve our data pipeline reliability. 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 my team can focus on scaling our data systems to meet the needs of Pubmatic's rapidly growing business.</i>”</p> <p>ASHWIN PRAKASH PUBMATIC DATA ANALYTICS ENGINEERING LEADER</p>  |
|---|---|--|

*As of December 2020