Chartboost sees significant AWS savings with Pepperdata

Customer challenges:

- Poor visibility into cluster activity for maintenance and troubleshooting
- Lack of job regulation to ensure on time job completion
- High cost of running underutilized Amazon Web Services clusters

Results with Pepperdata:

- 31% average gain in overall cluster throughput
- 30% reduction in Amazon Web Services (AWS) nodes

Background

Chartboost, a San Francisco-based company, is the largest mobile-games only platform to increase revenue and discover new players. Chartboost's technology provides mobile game developers with free cross promotion, a free developer-to-developer direct deals ad marketplace, the largest mobile games-only ad network with interstitial and mobile video ads, plus robust app analytics.

With over 700 million monthly active players interacting with 200,000 games that have integrated with Chartboost's SDK, the company ingests a high level of data each day. To process this, Chartboost originally deployed 33 nodes of varying sizes in Amazon Web Service's (AWS) Elastic Compute Cloud (EC2), mostly d2.8xlarge and r3.8xlarge instances.

Key Challenges

Prior to installing Pepperdata software for Hadoop, Chartboost was faced with the challenge of managing explosive growth in its data requirements, not just for servicing the analytics and ETL jobs supporting its core products, but for internal business intelligence purposes as well. "We needed raw processing power, and various Chartboost teams were going off and spinning up their own AWS clusters," says David Clubb, Senior Data Engineer at Chartboost and member of the data team responsible for data-related initiatives at the firm.

This was a costly approach to handling growth, because these ad hoc clusters would often sit idle when teams weren't directly accessing them. In addition, there was no way to view granular health and status metrics of all the clusters from a single, centralized dashboard for effective monitoring and troubleshooting. Finally, although Chartboost was using YARN's Fair Scheduler to spin up jobs, it was impossible to regulate them once they were running to guarantee high priority workloads could complete on time if additional tasks were subsequently scheduled.

Why Pepperdata: AWS cost savings, control, and deep visiblity

To address these challenges, Chartboost needed to deploy a single cost-effective, multi-tenant cluster that could run several workloads simultaneously — without causing performance

_____

> "We immediately saw performance gains — what's better than that? We're very happy."

_____

degradation for its most critical jobs. For this to even be possible, Chartboost needed a tool capable of making thousands of decisions per second to handle real-time resource contention

from multiple workloads.

Chartboost installed Pepperdata software for Hadoop on their primary AWS cluster in August of 2015. Because Pepperdata software can dynamically assign more work to underutilized containers that have been statically allocated by YARN, Chartboost was able to improve throughput on its existing infrastructure by 31 percent. Ultimately, this allowed the company to reduce its total AWS node count by 30 percent. Chartboost now deploys a single 22-node cluster using Amazon Web Services (AWS) d2.8xlarge instances, each node comprising 244GB memory and 36 cores.

"The throughput improvement is worth the cost of Pepperdata for us," says Clubb, "We installed on the production cluster. We immediately saw performance gains — what's better than that? It gave us a performance boost and we could measure that easily. We're very happy."

Chartboost can now ensure mission critical jobs complete on time, no matter what other workloads simultaneously contend for resources, by configuring Pepperdata policies. These can be set to prioritize specific users, jobs, or queues at Chartboost's discretion. The best part is that configuration is a set-it-and-forget-it feature of Pepperdata software: enforcement happens dynamically when contention is sensed as priority jobs are running. Today, Chartboost successfully runs anywhere from 20 to 30 simultaneous applications on its primary cluster, which

translates into 1,000 concurrent tasks, on average.

Finally, deep visibility into over 200 metrics captured second-by-second (down to the task- and user-level) across every node in the cluster provide actionable insights into cluster activity. "Being able to break down CPU usage by user, job, and queue is really useful," says Clubb. "Once we installed Pepperdata software, we were able to see that we had a single, large job that was consuming significant resources — it was causing nearly 100 percent CPU utilization multiple times a day." Before Pepperdata, Chartboost had been able to see that CPU usage was high but the tools used couldn't drill down enough to display which specific job was causing issues.

In short, installing Pepperdata software has allowed Chartboost to significantly reduce its operating expenses and AWS node count, guarantee mission critical jobs complete on time, and rapidly troubleshoot cluster issues thanks to deep visibility across all nodes.

About Pepperdata

Pepperdata is the only solution for Hadoop and Spark that enables predictable performance and true SLA enforcement for mission critical jobs, applications, or users. With Pepperdata, you can run multiple, concurrent workloads on a single cluster with consistent, reliable performance.

Pepperdata software is trusted on over 10,000 production nodes in firms of all sizes — including 3 Fortune 50 companies.