clearscale

aws partner network

Premier Consulting Partner

Complete Infrastructure for Batch Processing System and Broad Network of Content Websites Migrated from Joyent to AWS Enabled by Chef

MOB SOC media

Summary

MobSoc runs a publishing network with over 57 responsive media sites across entertainment, sports, lifestyle, and business/technology categories. Their proprietary Social Rank algorithm is driven by a batch processing system that collects and analyzes social metrics from Facebook, Twitter, LinkedIn, Pinterest, Google Plus, and YouTube.

ClearScale performed an audit, architecture design, and migration while automating everything with Chef. ClearScale architected and built out MobSoc's network of responsive sites from the ground up and completely re-architected and optimized MobSoc's batch system for AWS.

Executive Summary

MobSoc Media enables brands to reach the right vertical audiences with relevant branded content. Their proprietary technology, Social Rank, enables users to easily discover the most popular and trending content in areas they care about and to engage with like-minded people.

For their customers, who are brand marketers and advertisers, MobSoc offers precise targeting of relevant audiences in the mobile and social sphere. As a result, MobSoc delivers some of the highest engagement rates in the industry — up to seven times higher than the category average.

The Challenge

MobSoc Media enables brands to reach the right vertical audiences with relevant branded content.

MobSoc had been using Joyent to host their applications, batch systems, FanNewscast website, and core APIs since 2008. But as MobSoc grew, brought on more clients, and further developed their article import, social metric, and indexing applications, Joyent was no longer a fit.

The team needed better tools, managed services, and a greater depth of cloud offerings than what Joyent had available. Building out those services themselves would not have been cost-effective.

"In Joyent, we would have had to build and manage most of these services ourselves, which would pull valuable time and resources away from improving our product and refining the user experience," said Clayton Bolz, VP of Engineering at MobSoc Media.

> "ClearScale has been very valuable in our efforts to operationalize and scale our SaaS platform. They have strong technical resources that accelerate our delivery capability and bring valuable experience to our development team."
>
> **Clayton Bolz,** VP of Engineering, MobSoc Media

The ClearScale Solution

MobSoc chose to work with ClearScale to architect, implement, improve, and migrate their architecture from Joyent to AWS.

Planning the Migration to AWS

ClearScale worked with MobSoc to understand their current system on Joyent and created an architecture audit document of the existing deployment and proposed a new deployment architecture on AWS.

MobSoc had a complex deployment that was responsible for executing batch processing jobs to import, process, extract, store, and index article data and social metrics. Plus, MobSoc wanted to build out a new content network of dozens of responsive websites (WordPress multi-site).

The teams decided that the project would be broken out into two phases:

Phase 1: Build out new WordPress deployments on AWS

The goal of the first phase was to prove out a new architecture on AWS. In this phase, ClearScale focused on architecting and building the network of responsive sites from the ground up on AWS.

ClearScale set up all services, roles, and groups on AWS including custom AMIs. They developed Chef cookbooks and CloudFormation templates. Later they deployed app servers, database clusters, and other components for the production, staging, and development environments.

As part of the testing process of this phase, ClearScale launched the production, staging, and development environments and used JMeter and several testing scenarios to tested HA and scalability of the databases and application array.

With phase one complete, MobSoc had in hand a fully-functioning AWS deployment to manage their network of content websites.

Phase 2: Re-architect for improved batch processing and automation

For the second phase, instead of simply mapping over the Joyent deployment to equivalent services on AWS, MobSoc wanted to improve batch processing for their application and update their architecture with better automation capabilities. This included migration of their batch systems, FanNewscast website, and core APIs from Joyent.

MobSoc uses a sophisticated batch processing system to take in data from RSS feeds and update social metrics from Facebook, Twitter, LinkedIn, Pinterest, Google Plus, and YouTube. Their customers are able to use this information to precisely target relevant audiences. MobSoc wanted to be able to import new articles as soon as possible, and to optimize how they checked for updates to social metrics, such as new likes on articles.

ClearScale re-architected batch jobs to use Gearman, which would allow MobSoc take advantage of the best horizontal-scaling batch processing techniques. The Gearman content processing cluster consists of the following parts:

- **Gearman Server** is a generic application framework that holds the job queue and farms out tasks to workers.
- **Batch Controllers** submit tasks to Gearman server for processing.
- **Batch Workers** fetch tasks from Gearman Server and execute them.
- **Tasks** are small applications in PHP that execute work.

ClearScale created content processing scripts that are responsible for initiating tasks to import new RSS content and social metric updates. Each task is a batch of articles to fetch from RSS, Facebook, Twitter, or one of the other social sites. There are also tasks for information storage and results processing. Parameters can be set to determine iteration frequency and processing priority. All jobs provide execution statistics to a central place for monitoring, profiling, and server capacity planning.

Gearman is connected to Memcached, a memory caching system that helps speed up web applications by alleviating database load. Data is stored in Amazon RDS. Chef is used to automate the creation of server, controller, and worker instances.

Migration from Joyent to AWS

After final testing and optimizations were complete, ClearScale performed a mock cutover to confirm that all configurations, instances, and security policies were running correctly.

With the mock cutover a success, the new production, staging, and development environments were officially deployed on AWS, DNS was updated, and the migration from Joyent to AWS was successful.

New Cloud Infrastructure on AWS

MobSoc now runs their full production, staging, and development environments on Amazon Web Services.

New Amazon EC2 instances are automatically deployed via AWS CloudFormation that has been configured with auto-scaling policies.

A scalable Amazon RDS deployment is used to store article data and social metrics coming in from the batch processing jobs.

MobSoc now uses AWS Elastic Load Balancer instead of Zeus Load Balancer to automatically balance traffic across servers, and Elasticache is used to improve the performance of their web applications.

MobSoc also uses Amazon Route 53 to manage DNS across their deployment.

Infrastructure Automation with AWS CloudFormation and Chef

Management of MobSoc's production, staging, and development environments on AWS is made easier with automation.

Creation of new servers is automated with Amazon CloudFormation which creates new application, image, and Sphinx Search server instances that are part of an auto-scaling array. The array is configured with autoscaling policies that monitor conditions such as CPU and memory utilization. CloudFormation eliminates the need to manually create servers.

Chef is used to automate server configuration and system settings. Each server type running in MobSoc's deployment has its own Chef cookbook with a set of recipes that define precise configurations and settings, including installing and configuring Apache, Wordpress, MySQL, PHPMyAdmin, and Memcached. Additional Chef cookbooks create user roles and permissions.

The batch processing Gearman servers have their own set of Chef cookbooks and recipes that download, install, and build Gearman servers, controllers, and workers as well as configure the servers to prepare them to receive and execute tasks. The complexity of the sophisticated batch processing system is now more scalable and customizable and ultimately more simple due to the infrastructure automation made possible by Chef.

High-Performance Indexing with Sphinx Search Server

MobSoc's batch processing system generates a massive amount of data, so they need a powerful system to index that data and make it available for search.

MobSoc uses Sphinx Search Server for batch indexing and searching new articles. MobSoc wanted to organize the indices by publication date, with shard1 for news published in the last three months, shard2 for news in the last three to six months, etc. Originally, Sphinx had been configured such that the application had to talk to shard1 to be able to access data from other shards. If shard1 went down, then there would be downtime for the whole application. To mitigate this, MobSoc had to launch two instances of shard1, a master and backup. However, this configuration was not ideal.

ClearScale re-architected the Sphinx configuration so that applications servers can talk to any instance in the Sphinx cluster. The Sphinx installation can access remote indices that are configured on each web server. Sphinx instances are configured with a Chef cookbook, and new Sphinx instances can be deployed by CloudFormation if auto-scaling policies are met.

Enterprise-Grade Monitoring with Zabbix

Because of the complexity of MobSoc's deployment, ClearScale set up Zabbix to monitor the health of MobSoc's deployment at all times and to immediately alert the team when there might be a problem. Zabbix is a critical tool and is particularly suited for enterprise-grade monitoring needs and complex infrastructures.

ClearScale created monitoring templates to monitor every service including ELB, RDS, and the application and image servers.

Several monitors are needed for Sphinx and Gearman. Zabbix looks for situations where there are too many jobs in the queue or when the Gearman server has too long of a response time. Zabbix also issues warnings and alerts when Gearman worker and controller processes do not match configuration, and when workers and controllers encounter errors. For Sphinx, Zabbix is listening for situations where Sphinx servers might be unavailable or inoperable and monitors for conditions like number of running processes, whether certain files exist, and the age of those files.

The Benefits

"ClearScale has been very valuable in our efforts to operationalize and scale our SaaS platform. They have strong technical resources that accelerate our delivery capability and bring valuable experience to our development team," said Clayton Bolz, VP of Engineering at MobSoc Media.

The new deployment on AWS is now much more powerful and scalable than the deployment on Joyent. The re-architected deployment now supports increased new users and enables growth. The new batch processing architecture supports faster indexing and optimized job processing.

On the development and managed services side, ClearScale supported the MobSoc team by enabling faster delivery and supplemented MobSoc's team with cloud experience.