Modern Analytics for ECommerce
Hadoop and Spark on AWS

Challenge An American multinational Fortune 100 seller of apparel, footwear, equipment and accessories needed to migrate its on-premises databases to Hadoop in the cloud, and also needed Spark for predictive analytics.

Solution Cloudwick planned and engineered a large Cloudera EDW migration from the client's data center to AWS to leverage the cloud for cost savings, agility and flexibility. This resulted in higher performance for the company's e-commerce site and modern BI analytics using Hadoop and Spark.

Benefit
- Cloudwick is a certified partner for both AWS and Cloudera with proven experience migrating production systems.
- Hadoop on AWS provides the client with the infrastructure, agility and flexibility required for developing modern analytic platforms and services that have rapidly changing system requirements.
- The company was able to change infrastructure within minutes to support new real-time predictive analytics and deliver improved inventory management, customer experience, marketing effectiveness and more.

Cloud Transformation

An American multinational Fortune 100 corporation, a designer and seller of apparel, footwear, equipment and accessories with over 40,000 employees globally, needed to migrate its Hadoop platform from its datacenter to AWS, and it needed to add predictive analytics capabilities The company chose Amazon Web Services (AWS) as its public cloud platform and Cloudwick as its big data open source solution provider to build, operate and manage the project.

Cloudwick, leading big data services and solutions provider to the Global 1000, migrated the on-premises Cloudera cluster to AWS to provide greater infrastructure agility, flexibility and global availability for the company's modern big data analytics platform. To ensure security and reduce risk, the bulk import from the Cloudera cluster was done via AWS Direct Connect, which provided private and secure connectivity, increased throughput, and a more reliable connection. AWS provides more reliability, agility and failover, as well as elasticity and flexibility for significant annual cost savings.

Analytics Advantage

In addition, Cloudwick integrated Spark into the Hadoop cluster to perform advanced analytics on the organization's clickstream, social media, location, customer purchase, and inventory data. Spark brings to the cluster a tool for accelerated queries, machine learning library, graph processing engine, and streaming analytics engine. Spark provides results quickly, allowing the company to reach more precise and accurate answers.

Cloudwick finished the migration in under four weeks, taking the company from pilot to production in just 18 days. Previously, the company had been unable to perform high-performance analytics, which meant it was at risk of falling behind the competition. Today, it can conduct real-time predictive analytics for inventory reporting, which means it is able to more cost-effectively manage its e-commerce inventory, improve customer service and marketing programs, enhance its online user experience and more.

Cloudwick
Big Data Success

To Become a Data-Driven Enterprise,
Contact: services@cloudwick.com
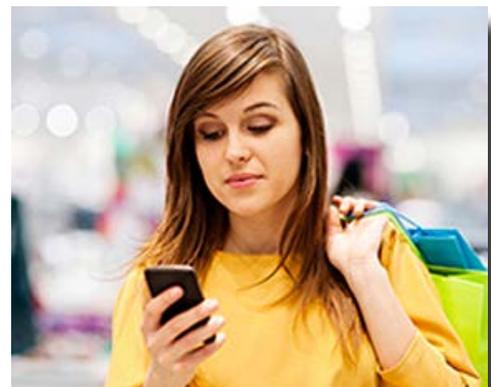
Data Variety

The retailer derives data from a variety of sources, including clickstreams and data analytics from its digital marketing platform provider to points of sale numbers at its brick and mortar locations and more. These data files, currently stored in Oracle, are compressed using S3 and loaded by Cloudwick into Hadoop using Sqoop on the 80-node AWS cluster. Once the data is loaded, there are 250 users within the company working on use cases that run analytics on the data using Hive, Impala and Spark. From this, additional line of business users, several thousand, use Tableau data visualization to generate reports on any number of business details to help the retailer improve its bottom line. Examples include:

Inventory and sales analytics: It runs analytics on a particular product line, examining data on the units sold, left in inventory, where they sold best, including online versus in-store, and more, helping it cut future production, shipping, staffing and other costs.

Marketing and advertising: The company tracks data to examine how many times a customer visits the web site based on IP address or site log-in to determine which platform - mobile versus desktop - produces the biggest sales, helping them more effectively determine where to spend advertising budgets.

User Experience: The customer satisfaction team is interested in studying data on how long it takes a shopper to successfully make a web purchase in order to ensure overall satisfaction. If a trend is noticed toward shopping cart abandonment, data is further examined to determine if the reasons are website or server issues so they can be immediately rectified.

New Solution Results

With Cloudwick's solution, the global ecommcerce company is running smoothly and efficiently with a cost-effective,

modern, cloud-based BI and analytics infrastructure, having eliminated a costly enterprise infrastructure. Marketing, user experience, inventory and other predictive analytics use cases are running smoothly for line of business leaders. With the move to AWS, the company can scale as needed and enjoy the cost savings with the flexibility and elasticity of the cloud. The company enjoys a flexible, agile, scalable and accelerated predictive analytics platform that provides it with precise answers quickly, gaining competitive advantage from its big data.

AWS Services Used

Elastic Cloud Compute to ensure the company has enough room for data but not to much (overpaying).

CloudWatch to monitor AWS resource utilization, application performance, and operational health.

Simple Storage Service (S3) to store & retrieve data.

Elastic MapReduce (EMR) for quickly and cost-effectively processing vast amounts of data.

Amazon Relational Database Service (RDS) to establish, operate, and scale the database in the cloud.

Direct Connect for private connectivity, increased throughput, and reliable connection for the migration.

Virtual Private Cloud (VPC) for a logically isolated section for launching resources in a virtual network.

Identity and Access Management (IAM) for managing users and user permissions.

CloudTrail to record API calls and deliver log files for security and compliance.

Route 53, a highly available and scalable cloud DNS web service, to connect user requests to infrastructures running in AWS and to route users to infrastructure outside of AWS.

Simple Queue Service (SQS) to transmit any volume of data, at any level of throughput, without losing messages or requiring other services.