A storage company migrates their on-premise Hadoop Cluster to AWS cloud

Executive Summary

Seagate selected AWS for its advanced analytic cloud platform and hired Cloudwick to perform the migration after soliciting multiple bids from four global system integrators.

Customer Challenge

Seagate's on-premises enterprise Hadoop cluster (EHC) consists of roughly 160 data nodes with 2 PB of data. The IBM BigInsights managed clusters were nearing the capacity and data center were to decommission in 5 months. Seagate wanted to migrate the workload to cloud.

Why did Seagate select Cloudwick as an AWS partner?

The Cloudwick team was introduced to Seagate through AWS sales team. Seagate made it clear from the beginning that they are taking bids from multiple vendors, including some that were already familiar with the project. Seagate's requirements for its Hadoop migration partner included the following; 1) partner had to be a certified AWS APN Data & Analytic Partner, 2) The partner had to have completed more than 25 enterprise production Hadoop engagements at Petabyte scale, 3) The Partner had to have successfully migrated Petabyte scale Hadoop clusters for Fortune 500 companies to Amazon, 4) The Partner had to have a Hadoop migration solution architect and engineering team that demonstrated expertise in large scale production Hadoop migrations, and 5) The Partner had to have a proven approach and methodology that was tested and proven. Cloudwick was one of four global GSI's that Seagate received RFP responses from, and they selected Cloudwick because it scored the highest in their review and demonstrated a clear knowledge and understanding of how to mitigate risk and deliver on time and within budget.

AWS and Cloudwick Solution

Amazon services used by Cloudwick in the migration include: S3, EMR, RDS, Snowball, Direct Connect, IAM, CloudFormation, CloudTrail and VPC.

Cloudwick started the migration with a business and technology workshop to identify key requirements, document the current state environment, identify the governance, compliance and security architectures, document the queries, jobs and their performance SLA's and build the RACI for managing the project amongst stakeholders. This migration required a considerable data migration plan that had to support business continuity and enterprise risk management as the services running on their Hadoop clusters was critical to their ongoing business.

About Seagate
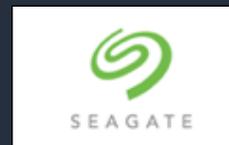
Seagate Technology PLC is an American data storage company. It was incorporated in 1978, as Shugart Technology. Since 2010, the company is incorporated in Dublin, Ireland, with operational headquarters in Cupertino, California, United States. Seagate has around 41,000 employees with 10.77 billion revenue in FY 2017.

Cloudwick met the aggressive four-month migration schedule and successfully migrated SeaGate's Big Data Analytics from Hortonworks to Amazon EMR saving SeaGate hundreds of thousands of dollars in software licensing too.

Next Steps

After the production workload runs for 3 months, Seagate will conduct a cost analysis on the queries on the cluster and go with reserved instances based on the analysis that will further reduce the client's spend on AWS. Also, there are opportunities for machine learning on the data available on S3.

About Cloudwick
Cloudwick is the leading provider of enterprise business and technology modernization services and solutions to the Global 1000 and helps enterprises gain competitive advantage from open source, data lake, big data, cloud and advanced analytics. Cloudwick is an advanced AWS partner with Big data and machine learning competency.

Cloudwick