Mactores & Seagate Case Study

The Challenge

Seagate Technology is a US-based data storage company with manufacturing facilities worldwide, generating vast amounts of manufacturing and testing data used by over 2,000 global team members, averaging 2.5M queries per day. With petabytes of data accumulated over 20 years and more being generated each day, it became imperative for them to have systems in place to ensure that the cost of collecting, storing, and processing data did not exceed the return on investment (ROI). The customer's on-premise big data stack was built on Oracle Data Warehouse for low latency query access and Hortonworks Hadoop Distribution with Apache Hive and Apache Tez as a big data storage and query processing. Their on-premise Factory Data Platform proved ineffective in terms of time and money for the large amounts of data being processed and queried by multiple users. In addition to these factors, end-users were dissatisfied with the performance of the queries that were taking a long time to complete, and sometimes these queries failed. Additionally, the maintenance and upgrades of these systems were compounding the problem. Seagate decided that they needed to partner with an organization that could provide the guidance and expertise needed to transform their on-premise big data stack to scalable data platform on AWS.

Why AWS

High availability enabled across the AWS global infrastructure was a key differentiator, as their manufacturing data is captured globally across North America, APAC, and Latin America. Amazon EMR is a highly scalable big data platform that supports open-source tools such as Apache Spark, Apache Hive, Apache HBase, Apache Flink, Apache Hudi (Incubating), and Presto. When you combine it with the dynamic scalability of Amazon Elastic Compute Cloud (Amazon EC2) and scalable storage of Amazon Simple Storage Service (Amazon S3), you get the elasticity to run petabyte-scale analysis for a fraction of the cost of traditional on-premises clusters. .

SEAGATE

About Seagate Technology PLC

Seagate engages in the provision of electronic data storage technologies and solutions. Its products and services include network-attached storage, high-performance computing, data protection appliances, internal hard drives, backup and recovery services, flash storage, and related solutions.

Why Mactores

After deciding on AWS, Seagate worked with a couple of different AWS GSI partners with expertise in Big Data and Analytics. However, they ran into challenges prompting them to interview other companies. Because Mactores had performed two successful deployments within Seagate for different divisions, and because of their AWS Big Data and Analytics Competency, the CIO reached out to Mactores to get involved. Mactores was asked to evaluate and deliver an alternative Data Platform and worked closely with the Seagate team to identify a path forward to solve the challenge of processing petabytes of data with consistent performance, lower query processing time, lower TCO and scalability required to support the 2,000 daily users.

Mactores & Seagate Case Study

The Process

At the beginning of the engagement, Mactores deployed their HEXA Audit®, which is a comprehensive Big-Data Assessment process which included the in-depth analysis of the customer's Big-Data requirements and were based on the following factors:

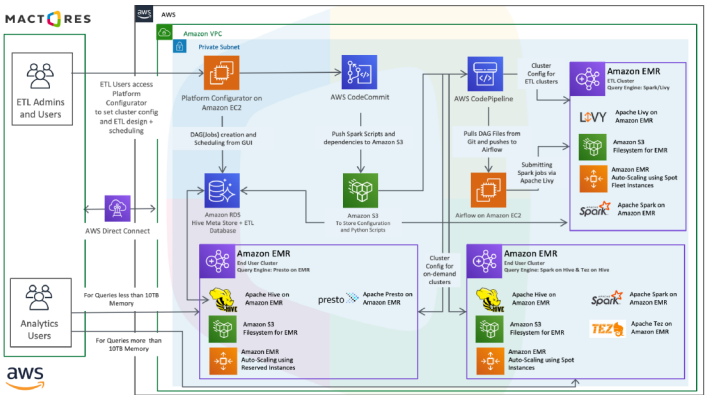
1. End-User Tools: Mactores needed to know what tools the end-users were currently using to query the data.
2. Analysis of long-running queries: They needed to know what queries were being used by the user and if there were any custom functions being used in the queries that may need custom solutions for migration.
3. Security and Access Control: Concerning security, Mactores had to collaborate with the information security team to understand the access control policies of the customer.
4. Storage: It was essential to understand the size of the data, the storage technology, and the feasibility of the migration to the new storage solution.

Benefits

The solution designed and deployed by Mactores enabled Seagate to continue processing petabytes of data at scale with significantly lower TCO. The migration to Amazon EMR Presto SQL resulted in annual savings of $1.4 Million, a 60% cost savings, and a 20x performance improvement. The Seagate Data Analytics and Sciences team can now scale up and scale down any number of jobs based on the budget and importance of the specific job. Additionally, a lot of CPU time was saved, and the End Users could get their Query Results in minutes rather than waiting for hours, thus improving the overall efficiency of the Amazon EMR Cluster and Business Operations. Seagate has also benefited from Mactores' accessibility and willingness to help on an ongoing basis, including ideation for other challenges.

The Solution

Mactores modernized the Seagate Hadoop Platform from Apache Hive to Apache Spark and Presto to support dynamic data querying capabilities. Data was ingested to S3 buckets using Amazon SDK from Seagate's Factory Information Systems. Once the data was available in S3, Mactores developed the ETL framework, which would autodetect schema updates and versioning of metadata to process the raw files into ORC files for Spark and Presto. This fully automated and resilient ETL process, Datalake and Data Querying capabilities through Spark, Presto, and Amazon EMR, was used to support 2000+ users across multiple geographies. To make this solution resilient and repeatable for multiple solutions, Mactores built a platform configuration management system (PCMS) integrating Amazon CodeCommit, Amazon CodePipeline, CloudFormation Template, and Ansible Scripts. The PCMS provided Seagate the ability to deploy the solution repeatedly with variations in the configuration for their new Datalake projects and analytics requirements.



- **Learn More**
- **Register for an Immersion Day**

Next Steps

Since the beginning of their engagements, Seagate has been able to reach out to Mactores to brainstorm solutions to solve other critical business challenges by working with Mactores to help identify the concise problem statements and the path forward to solving. Seagate is currently exploring other ways to transform their business and future by further leveraging Cloud technologies.

Website - www.mactores.com
Email - consult@mactores.com