Enabling a global telecommunications company to optimize the use of their greatest asset: their data

Objective

A global telecommunications leader started their journey to cloud computing by adopting Microsoft Azure. The company wanted to move away from costly on-premises data centers and into the cloud, while at the same time taking advantage of cloud-scale analytics, where possible.

Challenge

The company's cloud program was initially very application-focused, and the data science team quickly realized the cloud migration needed to account for — and leverage — the value of their data being available in the cloud. The challenge centered around migrating tens of petabytes of live production data without disrupting mission-critical business applications and operations.

Solution

The company adopted a data-first approach and used Cirata Data Migrator to move over 20 petabytes of data without requiring any business disruption. The data was immediately available for data scientists to use Azure services for AI and machine learning. Data teams went from blocking nearly 138 million robocalls per month to blocking over 1 billion robocalls per month, as a result of using cloud-scale analytics.

Business Impact

- Original data migration timeline objective was cut by over 50%
- Data-first approach enabled data scientists to immediately begin new AI & ML development
- Early development of high-value AI models enabled enhanced fraud detection
- Able to block 1 billion robocalls per month, over 7.2 times more per year than before
- No disruption to existing production environment during entire migration process
- Able to save millions of dollars by decommissioning on-premises disaster recovery data center

a great leap forward

50%

faster data migration

Customer Profile

This case study is based on a diversified global leader in telecommunications, media and entertainment, and technology.

Objectives

The company had started their journey to cloud computing by adopting Microsoft Azure. They wanted to move to the cloud for several reasons:

- To reduce costs and gain operational efficiency
- To gain the benefits of cloud elasticity
- To modernize and transform their technology environment

> "We cut our entire cloud data migration timeline for moving 13 petabytes in half."
>
> *Vice President of Data and Analytics, Global Telecommunications Company*

Challenges

Initially, the company's cloud program was very application-focused. This is a large company with thousands of applications, and as they started to look at their application journey, one of their big realizations was that it wasn't so much about the applications as it was about the data. This was especially true for the data science office, so the organization started to think about a "datafirst" approach and finding a way to get the data to the cloud to start getting benefits early. They wanted to be able to quickly demonstrate the business benefits of accelerating their business and improving customer experiences, revenue, and customer retention, as well as achieving cost savings.

The challenge was regarding the volume of existing data (tens of petabytes) and the fact that new data was continuously being ingested. How could the company migrate this volume of data efficiently and without impacting current business operations?

The company looked at a variety of solutions, including data transfer devices, ETL tools, and open source tools such as DistCp based software. Each of these solutions weren't fit for purpose. With data transfer devices, the data needs to be copied onto multiple devices, the devices must be transported by truck to the cloud data center, and the data then must be copied from the devices onto the cloud storage. After realizing multiple trucks would be needed, the security issues associated with shipping the data via trucks became clear. They would either need to bring their production systems down to prevent changes from occurring during the transfer process or undertake custom code development to identify and migrate any new or changed data. They quickly discarded transfer devices as a viable approach. Using ETL or open-source software tools also had similar issues in handling changing data, and the company estimated it would be too costly to develop and maintain the custom solutions they would need to develop with those tools. A better alternative was required.

Solution

The company ended up selecting Cirata Data Migrator to enable the data-first approach they were looking for and to automate the Hadoop-to-Azure data migration process without requiring any business disruption.

Initially, a short production pilot was conducted using Data Migrator to transfer 100 TB of Hadoop data directly from the company's on-premises production environment to ADLS Gen2 storage. The pilot was performed over a weekend without the need for any custom development and without any impact to their production systems. And, the data was immediately available for use by Azure services.

The pilot was very successful. It showed the company that they could achieve their data-first strategy with Data Migrator. To perform the migration even faster, they decided to put in an order for additional network bandwidth. However, nothing prevented them from proceeding with the current available bandwidth. Their goal was to migrate about 1 PB per month and get the initial set of data migrated from their on-premises Hadoop cluster into Azure within 12 months.

Results

As they began the migration, they found that Data Migrator could enable them to achieve their 1 PB goal even faster than a month. When the additional network bandwidth arrived, the migration was already ahead of schedule. Data Migrator was able to leverage the additional bandwidth that was now available, and the overall timeline for migrating the initial 13 petabytes of data was cut by over 50%—and would have been cut by even more over the other legacy approaches. The company has since migrated over 22 PB in less than one year and it continues to use Data Migrator in a hybrid environment to transfer ongoing data changes from on-premises to the cloud. The company's Vice President of Data Analytics noted that "We cut our entire cloud data migration timeline for moving 13 petabytes in half."

By following a data-first approach, data became immediately available to the data scientists to not only begin modernizing existing workloads, but to start building new AI and machine learning models that were able to provide new business benefits very quickly. One area they were able to quickly enhance was with regard to fraud detection. Leveraging the data, cloud elasticity, and new AI capabilities, they are now able to quickly achieve savings by identifying robocalls in seconds where it previously took them days, resulting in 7.2 times more blocked robocalls per year than before.

The migration was achieved without requiring any production system downtime or disruption to their business. The existing production environment remained in use during the entire process, and any ongoing changes were replicated to the new cloud environment as they occurred. The company was able to avoid any big bang approaches where all workloads are moved at the same time, and the migration strategy provided the organization with time to modernize and optimize the workloads for the new cloud environment. This allowed the company to make sure applications ran optimally and to take advantage of new capabilities available to them.

The company was also able to achieve their objective of reducing costs and gaining operational efficiency by decommissioning their on-premises disaster recovery center, since the data residing in the cloud could now be used for this purpose. This decommissioning alone more than paid for the migration project.

Overall, the company was able to achieve a very fast ROI on the project by leveraging a data-first approach.

Results Summary

- Original data migration timeline objective was cut by over 50%
- Data-first approach with Data Migrator enabled data to become immediately available to data scientists
- Faster development of high-value AI models enabled enhanced fraud detection
  - Now able to identify robocalls in seconds, where it previously took days
  - More than 7.2 times more blocked robocalls per year
- Existing production environment remained in use during entire process (no business disruption)
- Able to take time to optimize existing workloads for new cloud environment and avoid big bang cutover
- Able to save millions of dollars by decommissioning the on-premises disaster recovery data center
- Achieved very fast ROI on the project by leveraging a datafirst approach

5000 Executive Parkway, Suite 270, San Ramon, CA 94583
US +1 877 (926-3472) EMEA +44 (0) 114 3039985
APAC +61 2 8211 0620 All other +1 925 380 1728

www.cirata.com