ARCITECTA®

CASE STUDY:

DIGITALGLOBE X AMAZON AWS

OVERVIEW

DigitalGlobe has been collecting imagery of Earth from space since its first satellite launched in 2001. Over the years, DigitalGlobe has built a world-class constellation, with increased collection capacity and resolution. With more than 7 billion square kilometers of images, DigitalGlobe's archive now totals 100 petabytes of storage, increasing by around 10 PB per year.

Given its relatively predictable usage curve and the expense of maintaining a 100 PB imagery archive on tape, DigitalGlobe approached Amazon to be the inaugural user of the AWS Snowmobile – a semi-truck filed with storage capacity equal to 1,250 Amazon Snowballs – an exabyte-scale data transfer service capable of moving extremely large amounts of data to AWS S3 and Glacier.

Mediaflux had been providing the core data management and workflow functions for the entirety of DigitalGlobe's 12,000 tapes since 2014; recalling and delivering any image in the archive to a customer within four hours – in 2017, this was done 4 million times. Arcitecta was tasked to work with Amazon to transfer all this data from the existing tape library to the Cloud – set to be Amazon's biggest data repository from a single client.

ARCITECTA®

DIGITALGLOBE X AMAZON AWS

THE CHALLENGE

Loading data from thousands of tapes into AWS Glacier is not what you'd call a simple file transfer! Whilst it is not uncommon for businesses to ship hard drives full of data to Amazon, sending 100 PB to individual hard drives or AWS Snowballs just isn't practical. Even when using the fastest available networks, transferring that many petabytes would take months, if not years, and would have left insufficient bandwidth for DigitalGlobe to carry on with its business as usual heavy data production.

THE SOLUTION

AWS Snowmobile acted like a giant hard drive that came to DigitalGlobe and was the right choice for the multi-petabyte problem. The first thing Arcitecta had to do was to enhance Mediaflux to support the management and archiving of data using AWS Glacier. That was easy, Mediaflux got a new storage interface, but data couldn't be transferred with this interface alone.

In a team effort DigitalGlobe, Amazon, and Arcitecta came up with a process to make data Glacier ready whilst moving it rapidly from DigitalGlobe's tape archive onto the Snowmobile. When the truck arrived at the AWS facility and Snowmobile unloaded the freight, it was Mediaflux that instantly managed petabytes of data as it was migrated from AWS S3 into Glacier; not just the images themselves, but all of the metadata that describes them.

To increase the speed capability for the transfer, Arcitecta created a plug-in for Amazon libraries to use cluster nodes to transfer the data. Initially, there were some issues. The Snowmobile could not handle the amount of data being pushed from the tape library to the trucks using Mediaflux – it was just too quick.

Amazon was able to re-implement some of its APIs and servers to handle what was possible, which has improved Amazon's platform as well; because now the Snowmobiles are a lot more resilient.

OUTCOME

Without Mediaflux working with Amazon, it would have taken DigitalGlobe an estimated two years to transfer 100 PB of data using available network links. The transfer with Mediaflux and AWS Snowmobile was a matter of weeks.

By modifying the platform to work with Amazon S3 and Glacier storage, Mediaflux saved DigitalGlobe a significant amount of time and money. DigitalGlobe has been managing its data through Mediaflux on AWS for nearly two years without stopping. No crashes, it just works.

ON REFLECTION

Working with DigitalGlobe and Amazon was a great experience. There were these three companies, massive ones like Amazon and DigitalGlobe, working hand in hand with a small technology company from Australia, to solve a huge data migration problem - and it was all seamless.

Using semi-trailer trucks to move data across the country might seem like overkill, but complex problems need big new ideas. Fortunately, as a collective, we had the technical capability and determination to pull it off.

For more information please contact talk@arcitecta.com
www.arcitecta.com
© Arcitecta Pty. Ltd. 2019