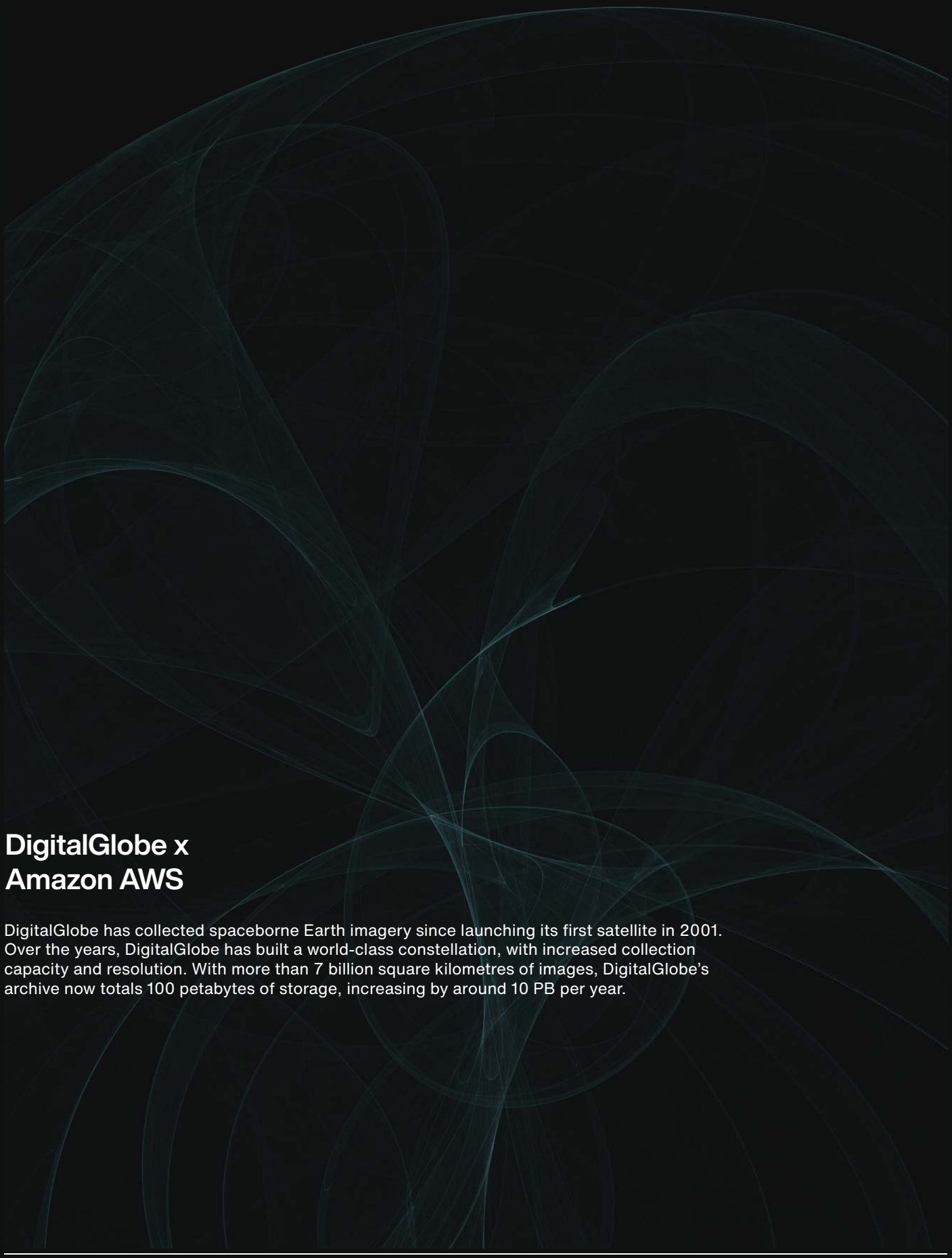
CASE STUDY
2017



DigitalGlobe x Amazon AWS

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

ARCITECTA
OPERATING SYSTEMS FOR META+DATA

www.arcitecta.com | talk@arcitecta.com

Overview continued

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-trailer truck filled 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 provided the core data management and workflow functions for the entirety of DigitalGlobe's12,000 tapes since 2014. This required recalling and delivering any image in the archive to a customer within four hours; 4 million times within 2017 alone. Arcitecta accepted the challenge to transfer the entirety of this data from the existing tape library to the Cloud, creating Amazon's biggest data repository from a single client.

The challenge

Loading data from thousands of tapes into AWS Glacier is not exactly a 'simple file transfer'! While 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 simply isn't practical. Even using the fastest available networks, transferring that many petabytes would take months, if not years, leaving insufficient bandwidth for DigitalGlobe to continue its normal data-heavy production.

The solution

AWS Snowmobile was an efficient means to solve the multi-petabyte problem, acting as a giant hard drive delivered to DigitalGlobe's door. Arcitecta first enhanced Mediaflux to support the data management and archiving using AWS Glacier. That was the easy part: Mediaflux got a new storage interface, but data couldn't be transferred using this interface alone.

A team effort between DigitalGlobe, Amazon, and Arcitecta produced an efficient process to make data Glacier-ready while rapidly moving it from DigitalGlobe's tape archive onto the Snowmobile. When the truck arrived at the AWS facility and Snowmobile unloaded the freight, Mediaflux instantly managed petabytes of data as it was migrated from AWS S3 into Glacier; not just the images, but all of their vital metadata as well.

To increase the transfer speed capability, Arcitecta created a plug-in for Amazon libraries to use cluster nodes to transfer the data. There were some initial issues: the Snowmobile could not handle the rate Mediaflux was pushing data from the tape library to the trucks – it was simply too quick!

Amazon was able to re-implement some of its APIs and servers to handle what was possible, which has

Outcome

Without Mediaflux working directly with Amazon, DigitalGlobe's 100 PB transfer using available network links would have required approximately two years. Instead, the transfer with Mediaflux and AWS Snowmobile was completed in a matter of weeks.

By modifying the platform to work with Amazon S3 and Glacier storage, Mediaflux saved DigitalGlobe significant time and money. DigitalGlobe has managed 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.