

# cloudy horizons

Lee Hibberd, Digital Preservation Officer, National Library of Scotland  
L.Hibberd@NLS.UK

DPC Procuring Digital Preservation Tuesday 21<sup>st</sup> March 2017



<https://graven.co.uk/portfolios/kelvin-hall-glasgow/>

# National Library of Scotland

*“..will be the guardian of the published and recorded memory of Scotland for current and future generations”*

+

*“By 2025 we will have a third of the Library’s holdings in digital format”*

≈ more digital content with the same staff and budget

Why aren't we using the cloud?

# **serious** **benefits**

- Keep a copy geographically separate
- Keep a copy on different technology/  
environment
- Keep a WORM or offline copy?
- Store it at lower cost?
- Undertake computation and health checks  
more quickly?

**BETA** This is a [beta service](#) – please send your feedback to [enquiries@digitalmarketplace.service.gov.uk](mailto:enquiries@digitalmarketplace.service.gov.uk)

[Digital Marketplace](#) > [Cloud technology and support](#) > Infrastructure as a Service

# Search results

## Choose a category

[All categories](#)

[Software as a Service](#)

[Platform as a Service](#)

[Infrastructure as a Service](#)

[Specialist Cloud Services](#)

## Keywords

**1840** results found in **Infrastructure as a Service**

### Amazon Elastic Block Store (Amazon EBS)

Smart421 Ltd

Amazon Elastic Block Store (Amazon EBS) provides block-level storage volumes for use with Amazon EC2 instances. They are highly available and reliable storage volumes that can be attached to any running EC2 instance in the same Availability Zone and are exposed as storage volumes that persist independently. [aws.amazon.com/ebs/](https://aws.amazon.com/ebs/)

Infrastructure as a Service   G-Cloud 7

1840

# let's play

Trials and technology tests

- > cloudy culture project
- > web API uploader
- > Oracle Cloud Storage

Cost modelling



100%  
FREE

# cloudy culture



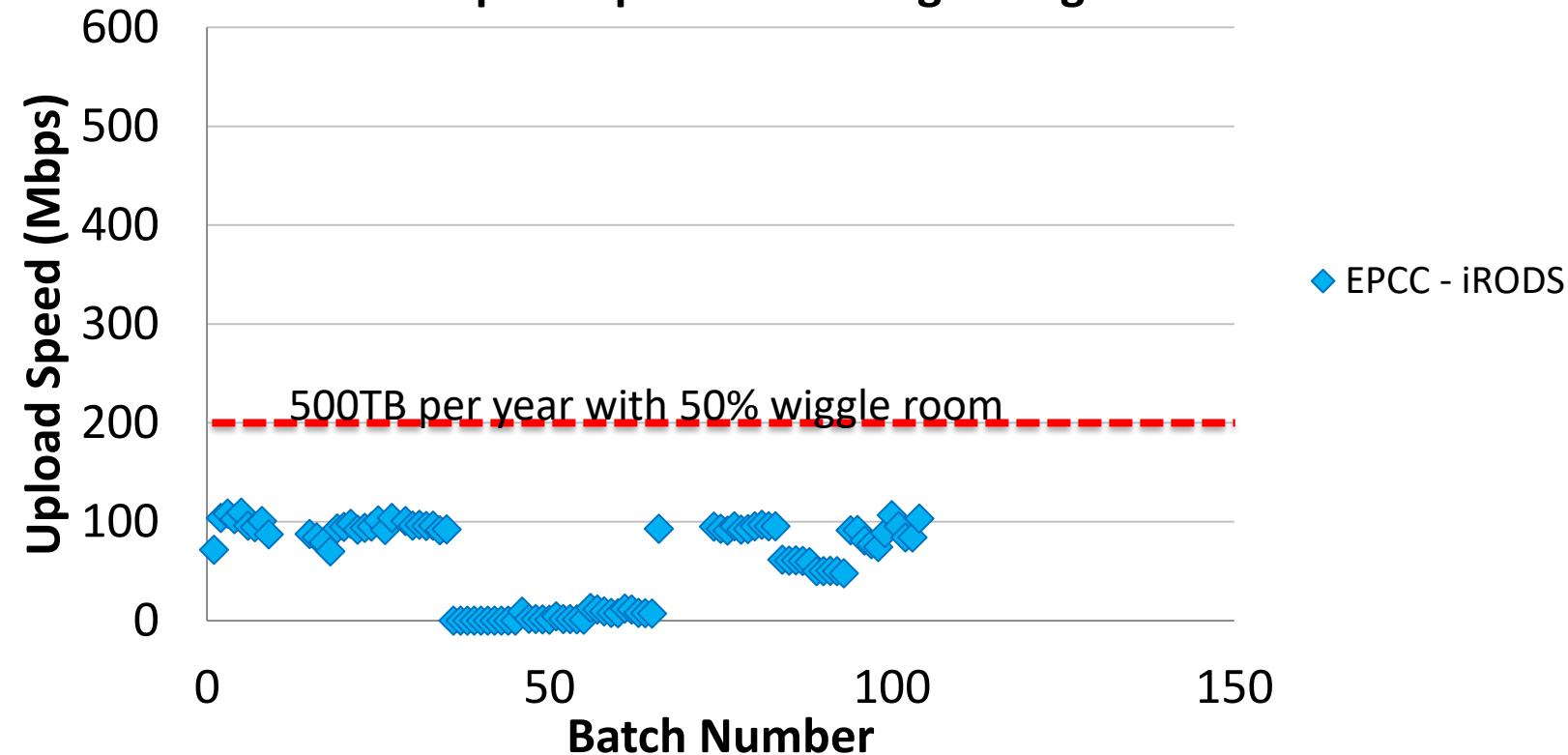
«cloud services to improve the preservation of digital cultural heritage»

<http://www.dpconline.org/blog>



# upload results

transfer speeds per batch using a range of services and tools



# 28<sup>th</sup> February 2017

BBC Sign in News Sport Weather iPlayer TV Radio

## NEWS

Home UK World Business Politics Tech Science Health Education Entertainment

### Technology

## Amazon data centre fault knocks websites offline temporarily



**Dave Lee**  
North America technology reporter

28 February 2017 | Technology

Share



Amazon has been promoting its cloud servers around the world

Several high-profile websites and services were knocked offline by a failure at one of Amazon's major US data centres.

# 19<sup>th</sup> August 2015

BBC Sign in News Sport Weather iPlayer TV Radio

## NEWS

Home UK World Business Politics Tech Science Health Education Entertainment

### Technology

## Google loses data as lightning strikes

19 August 2015 | Technology

Share



Lightning struck the power grid near Google's data centre four times in a row

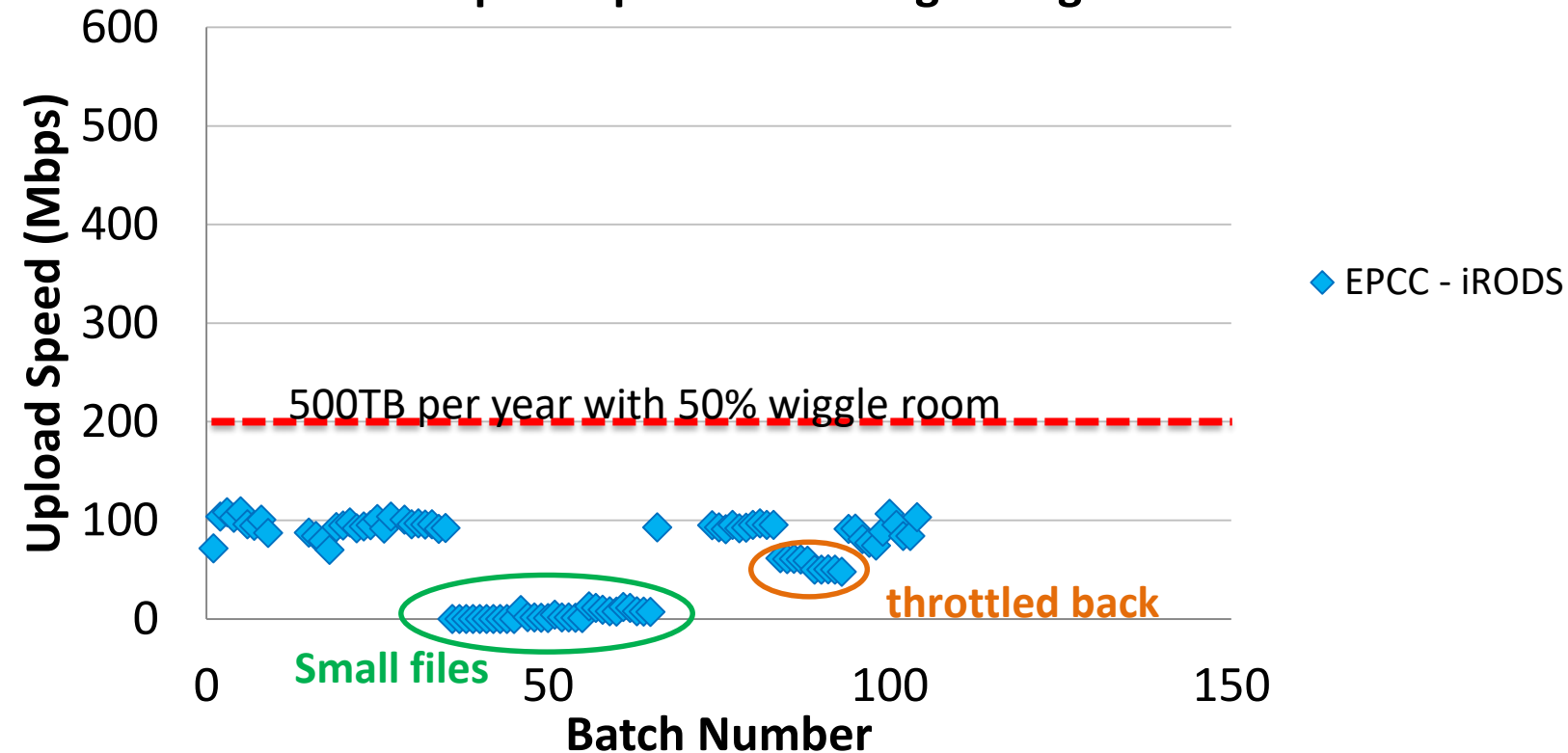
Google says data has been wiped from discs at one of its data centres in Belgium - after the local power grid was struck by lightning four times.

Some people have permanently lost access to the files on the affected disks as a result.

A number of disks damaged following the lightning strikes did, however, later become accessible.

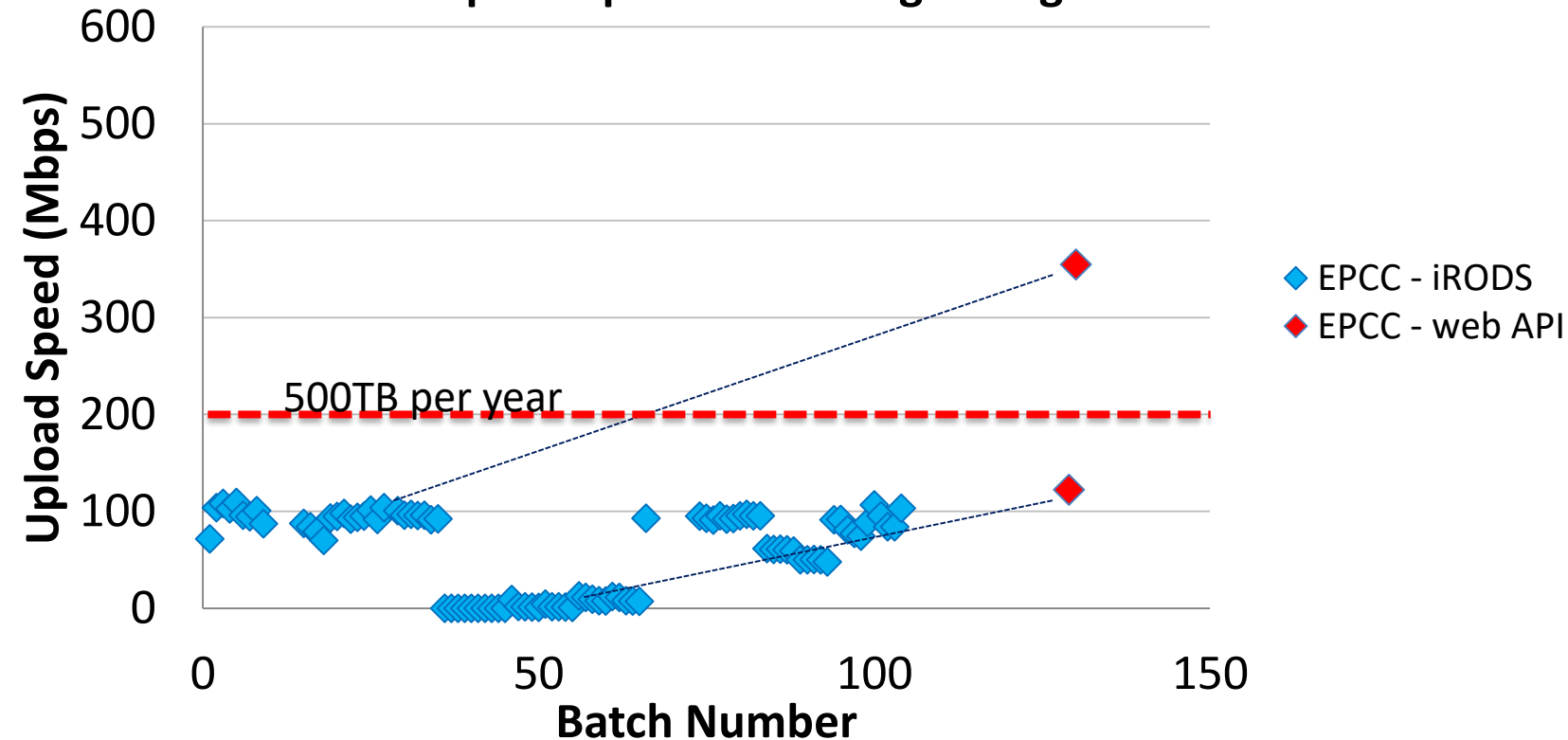
# upload results

transfer speeds per batch using a range of services and tools



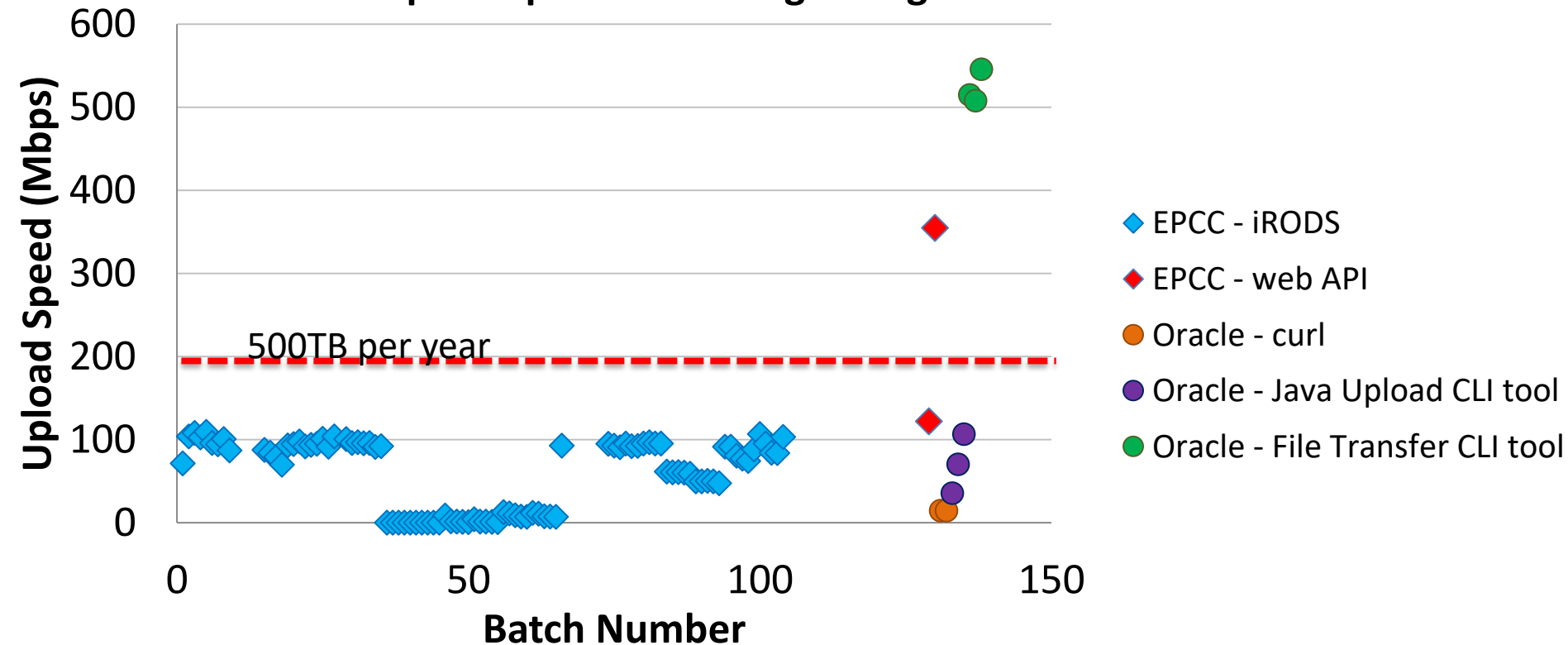
# upload results

transfer speeds per batch using a range of services and tools



# upload results

transfer speeds per batch using a range of services and tools

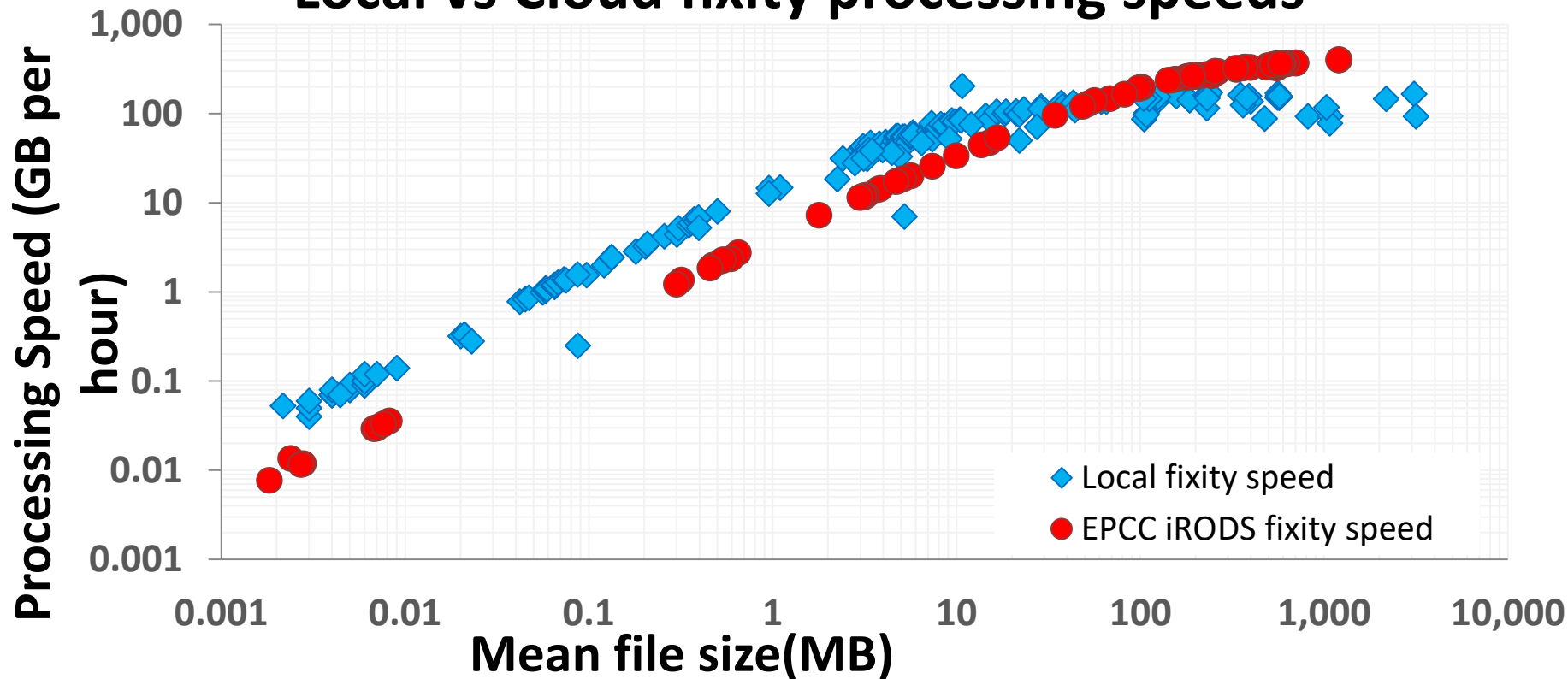


# consider **this**

- Use the right tools in the right way
- Use an appliance from the cloud provider to help manage the transfers in a simpler and optimised way
- Group data into larger data blocks to avoid the poor performance of uploading small files
- How are you sharing your network connection?
- Do you want to shuttle data?
- Allow time for disruptions

# fixity results

## Local vs Cloud fixity processing speeds







# cloud costs

## WHAT THEY TELL YOU

- Archive storage per GB per month
- Archive data retrieval per GB (3-5 hours)
- Archive data retrieval per GB (5-12 hours)
- Retrieval requests per 1000 requests (3-5 hours)
- Retrieval requests per 1000 requests (5-12 hours)
- Archive early deletion per GB (within 90 days of upload)
- Archive small read/writes (files <10MB) per 1000 requests
- Data transfer outbound first GB/month
- Data transfer outbound next 9.999 TB/month
- Data transfer outbound next 40 TB/month
- Data transfer next 100 TB/month
- Data transfer next 350 TB/month
- Put, copy or post requests per 1000 requests per month
- get and all other requests per 10000 requests per month
- delete requests
- listvaults, getjoboutput, delete and other Glacier requests
- shuttle charges

## WHAT YOU NEED TO KNOW

- How much data (GB)?
- What is the PUT error rate?
- What is the GET error rate?
- What proportion for fixity?
- What proportion for other downloads?
- How many files?
- How many copies of files?
- Annual service cost increase?
- Peak retrieval rate per hour
- Will you shuttle data on drives?
- Will you use dedicated PCs or a cloud service provided appliance?
- Will you use leased lines?
- Who will set up the tools?
- Who will hand-hold the transfers?
- Procurement costs?



Marko Karppinen [Follow](#)

CEO, Richie. We make app components for newspapers and magazines.

Jan 17, 2016 · 10 min read

## How I ended up paying \$150 for a single 60GB download from Amazon Glacier

In late 2012, I decided that it was time for my last

go. Between MacBook

Display, ours had suc

The 150-or-so CDs in

iTunes library, but a n

than AAC came along?

So I concluded that I n

thinking about tossing t

gigabytes. Certainly not

**Glacier data retrievals are priced based on the peak hourly retrieval capacity used within a calendar month. You implicitly and retroactively “provision” this capacity for the entire month by submitting retrieval requests. My single 60GB restore determined my data retrieval capacity, and hence price, for the month of January, with the following logic:**

- 60.8GB retrieved over 4 hours = a peak retrieval rate of 15.2GB per hour
- 15.2GB/hour at \$0.011/GB over the 744 hours in January = \$124.40
- **TAT for the total of \$154.25.**

**BETA** This is a [beta service](#) – please send your feedback to [enquiries@digitalmarketplace.service.gov.uk](mailto:enquiries@digitalmarketplace.service.gov.uk)

[Digital Marketplace](#) > [Cloud technology and support](#) > Infrastructure as a Service

# Search results

## Choose a category

[All categories](#)

[Software as a Service](#)

[Platform as a Service](#)

[Infrastructure as a Service](#)

[Specialist Cloud Services](#)

## Keywords

**1840** results found in **Infrastructure as a Service**

## Amazon Elastic Block Store (Amazon EBS)

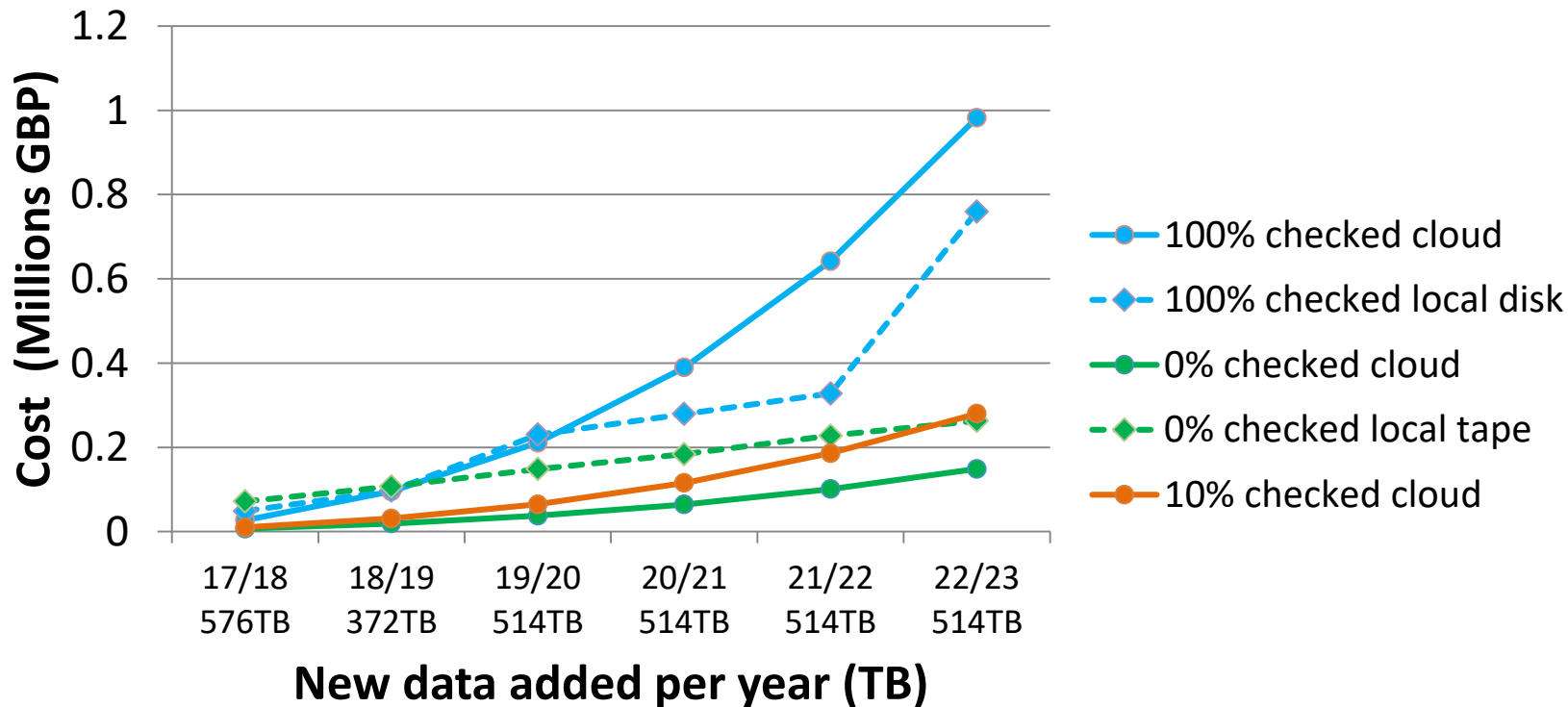
Smart421 Ltd

Amazon Elastic Block Store (Amazon EBS) provides block-level storage volumes for use with Amazon EC2 instances. They are highly available and reliable storage volumes that can be attached to any running EC2 instance in the same Availability Zone and are exposed as storage volumes that persist independently. [aws.amazon.com/ebs/](https://aws.amazon.com/ebs/)

Infrastructure as a Service   G-Cloud 7

# is the cloud cheaper?

cloud vs local cumulative storage costs



# conclusions for you

- No technical barriers for uploading and fixity checking content in the cloud
- Cheaper storage options limit the amount of fixity checking we can do
- Now need to gain confidence in the costs and performance of commercial services
- Increase the intelligence of our automated transfers
- Experiment with trials
- Understand the charges as best as you can
- Use websites to get costs and government Digital Marketplace
- Don't put all of your eggs in one basket – accept there is more risk while you build up experience of a particular service
- Speak to other cloud users