

Decisions! Decisions!

*Modelling the costs of a commercial
data archive service*



Agenda

- Context - a (very) brief introduction to Arkivum
- Questions to which we needed answers
- Key drivers in the model
- Some of our answers
- Conclusions



Arkivum in 60 seconds



SLA with 100% data integrity guaranteed



World-wide professional indemnity insurance



Long term contracts for enterprise data archiving



Fully automated and managed solution



Audited and certified to ISO27001



Data escrow, exit plan, no lock-in

Communities & markets



Higher Education



Communities & markets

TATE

*Heritage
&
Life Sciences*



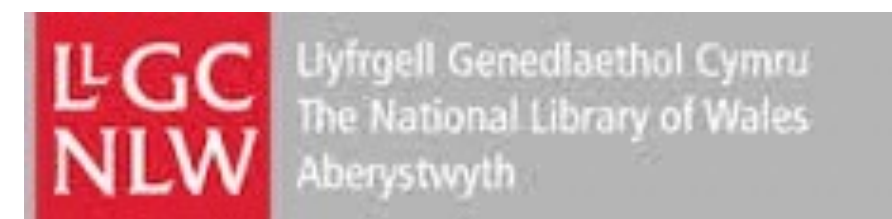
MoMA



Imanova
Centre for Imaging Sciences



Kuwait Investment
Office



Kew Gardens - Case Study

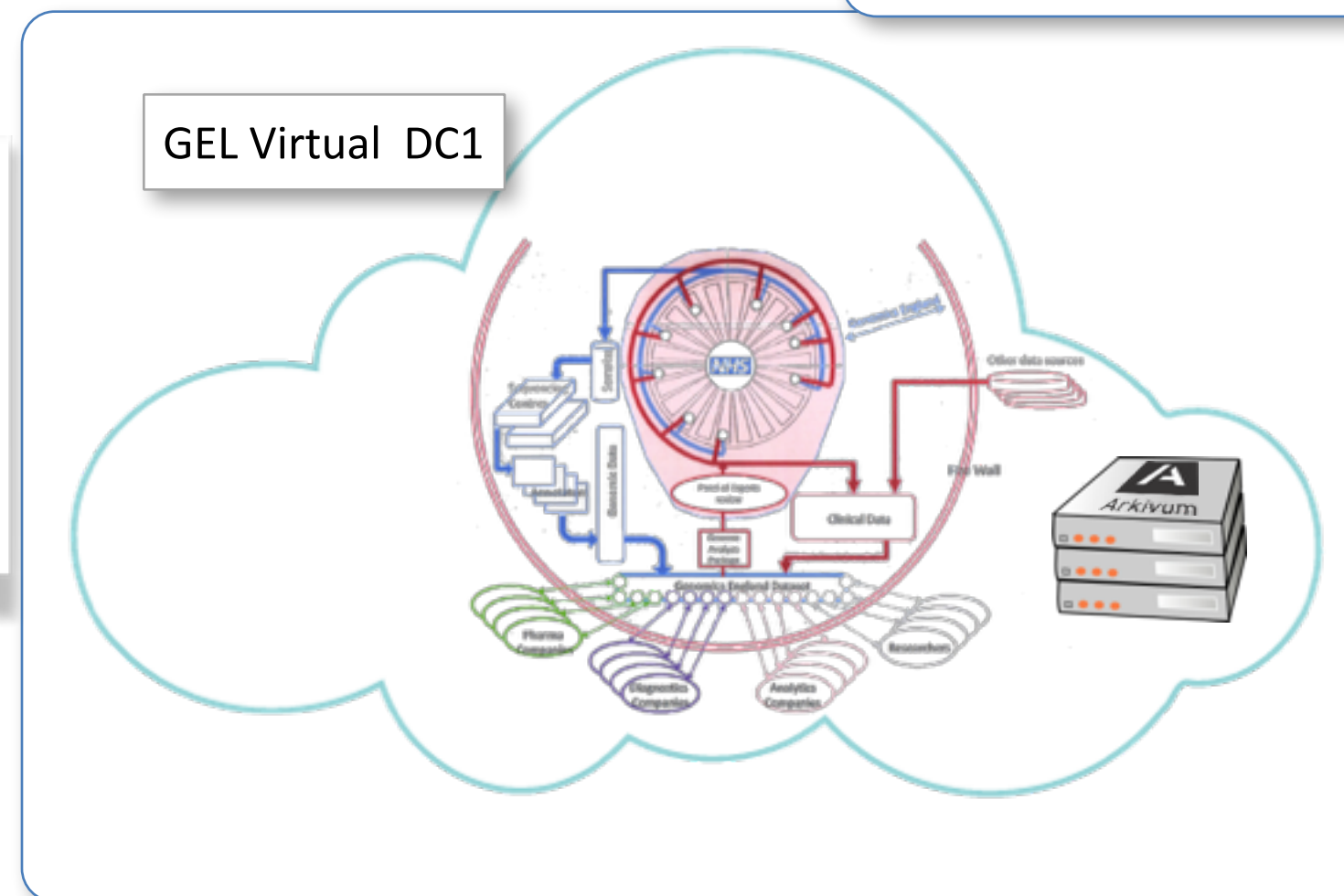
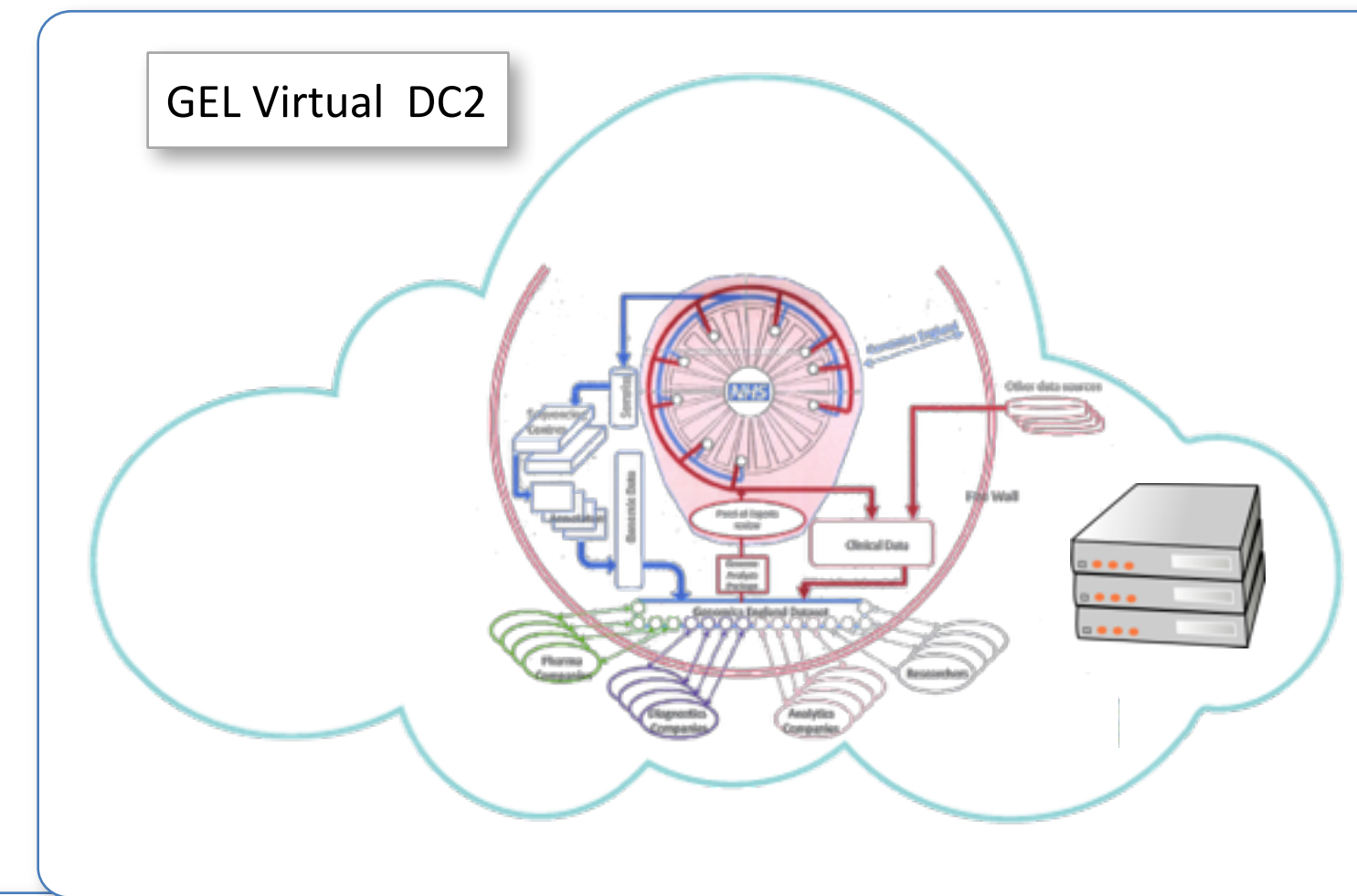
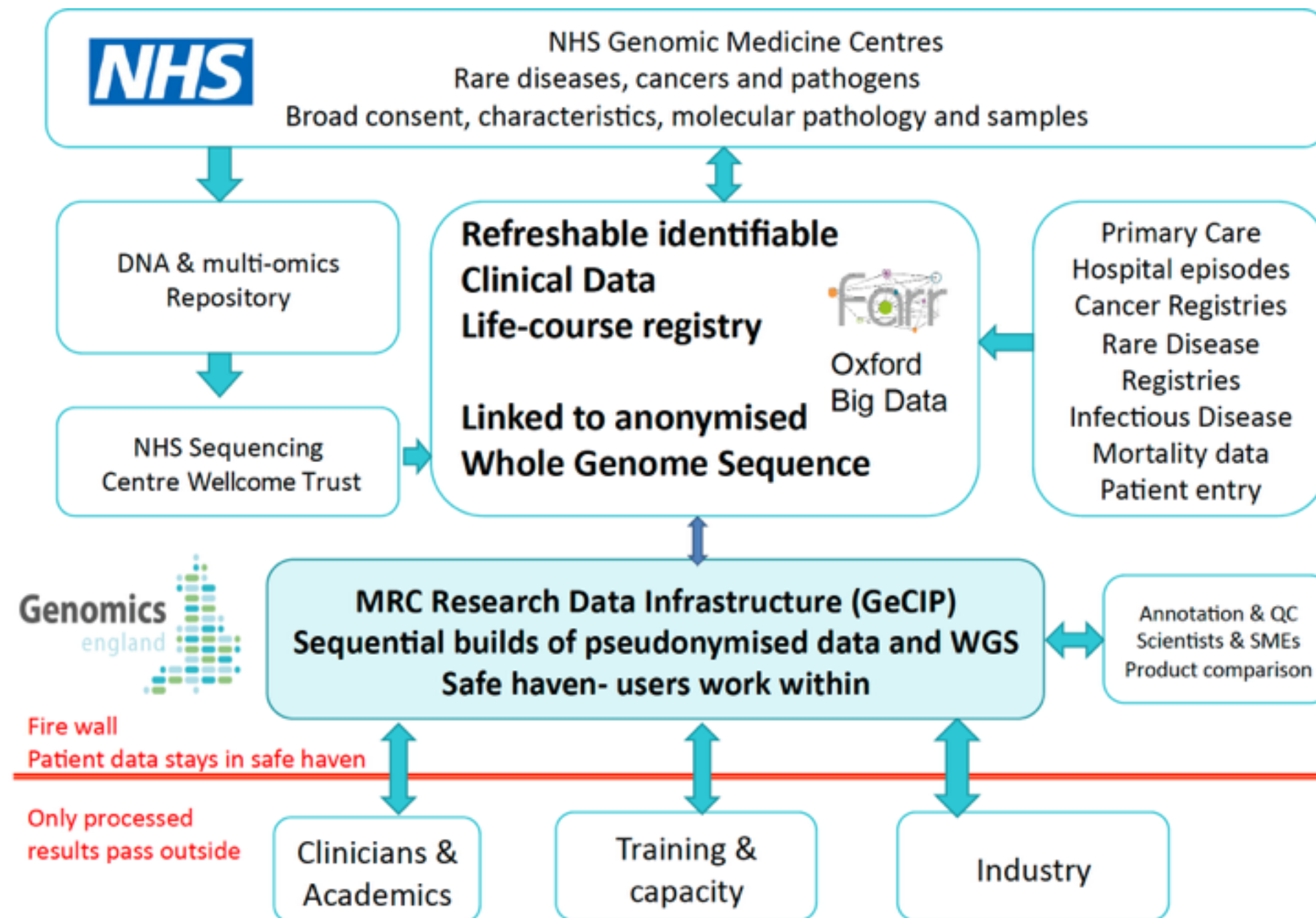
Since the beginning of IT at Kew in the mid-1980s, Kew staff have been transferring data from physical media to digital, creating new data about the collections, deriving data from experimentation and analysis, and, more recently, creating digital representations of the physical objects in the collections themselves.



Data growth:
5PB over 8 years

Genomics England

Genomics
england



GMCs & GeCIPs
eg
Oxford University Hospitals **NHS**
NHS Trust

**Data growth:
20PB by end of 2017**

Images courtesy of Genomics England

© Arkivum 2015/7



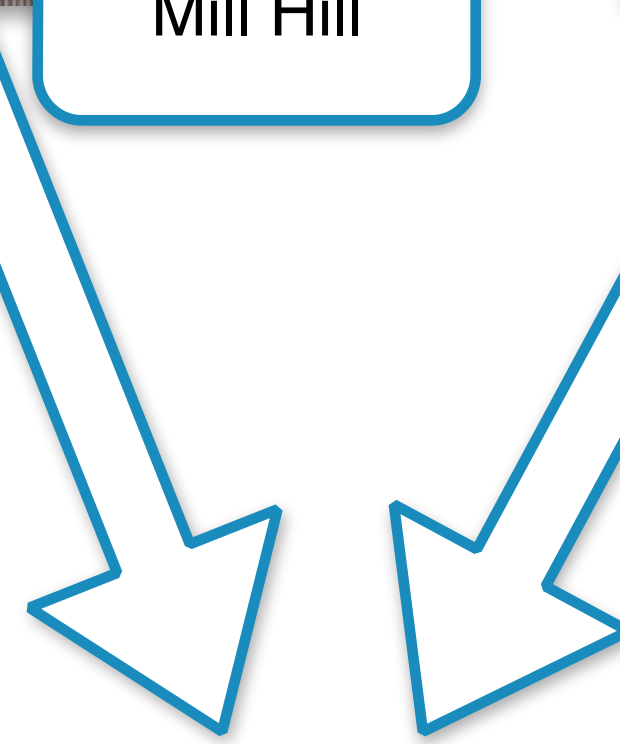
Francis Crick Institute - Lifeboat Project



Mill Hill



Lincoln's Inn Fields



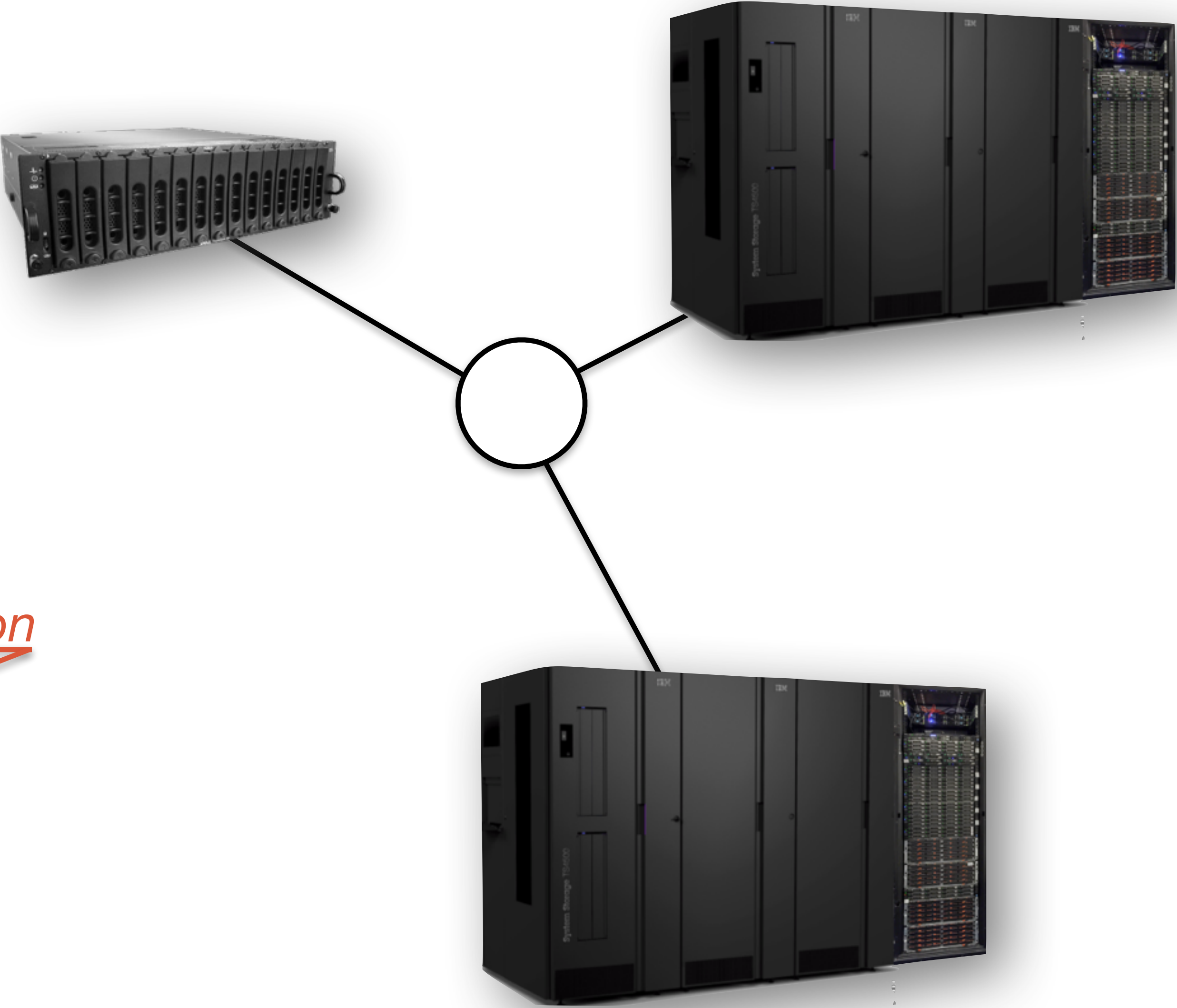
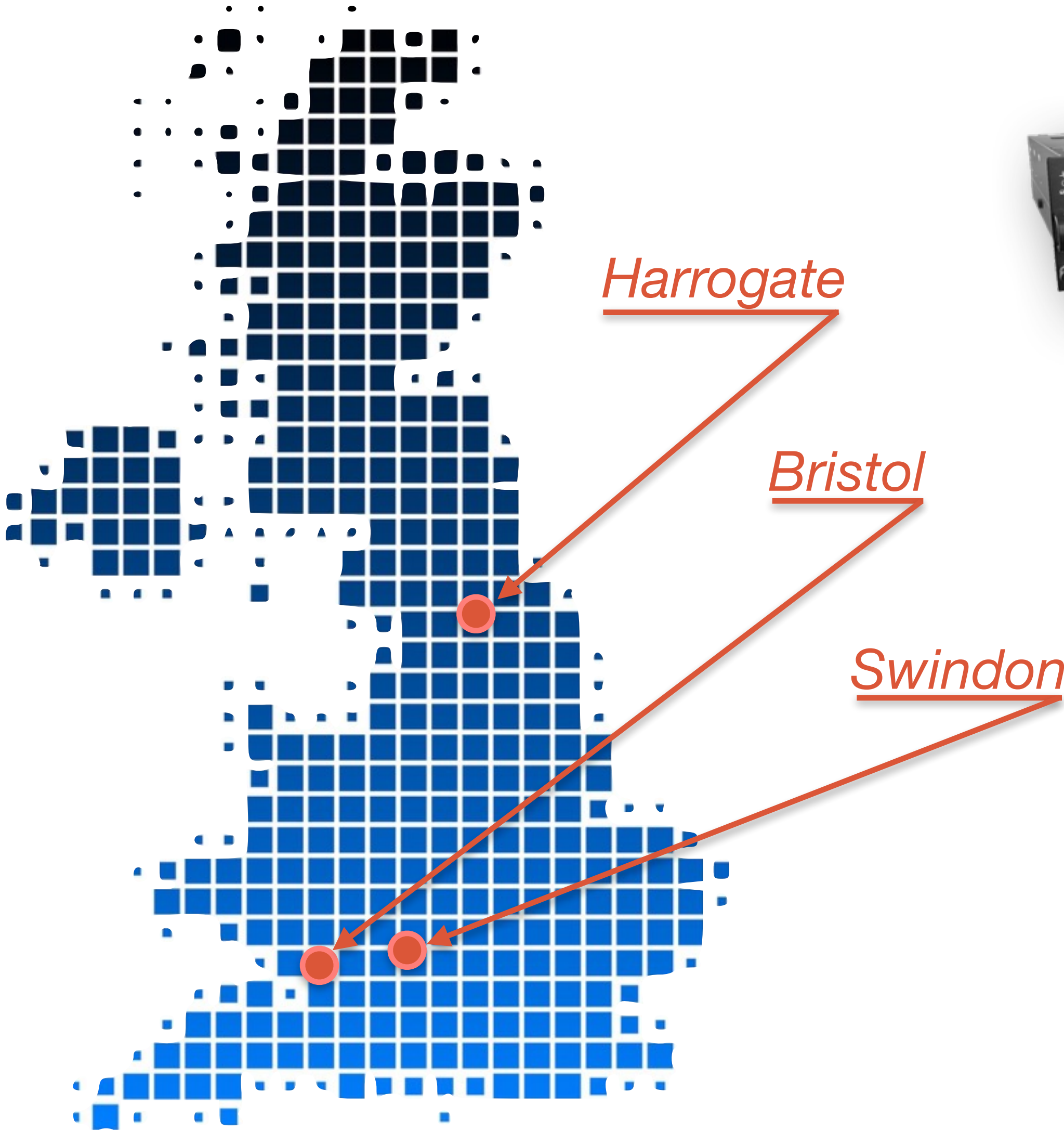
Brill Place



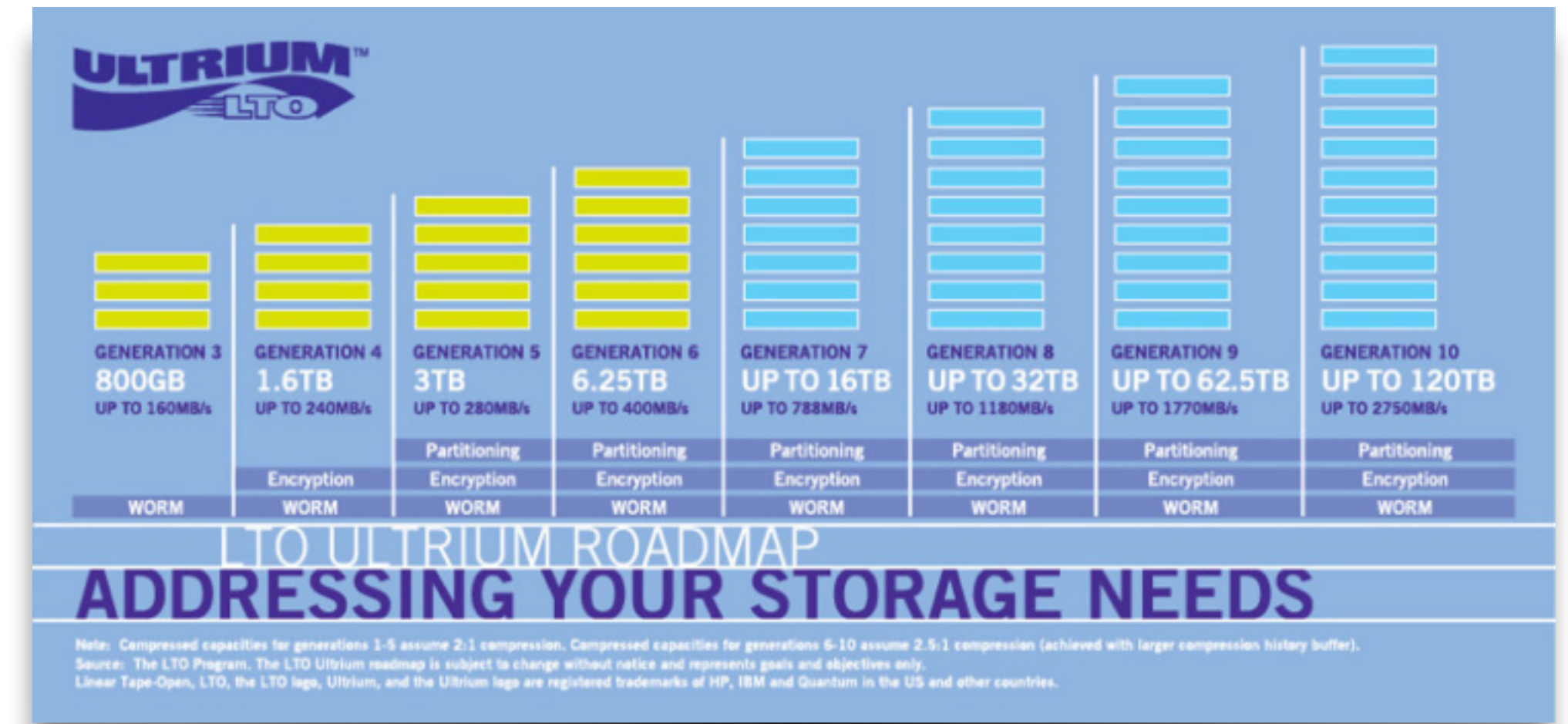
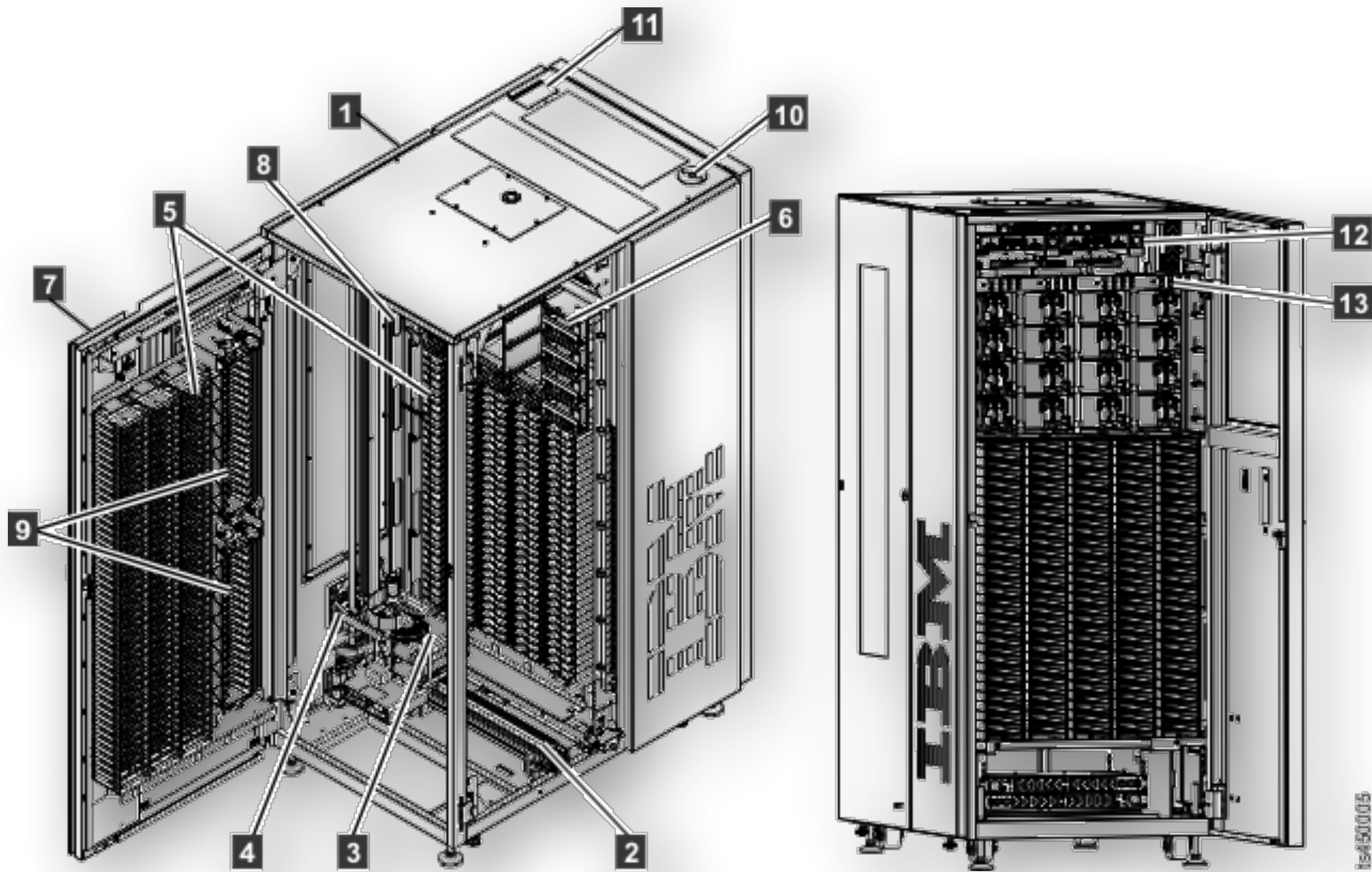
Consolidating 2.1PB of Isilon storage
Archiving critical data
Moving staff and equipment to Brill Place

Data growth:
2PB ingest over 3 months

Arkivum Infrastructure



Current bulk storage choice: IBM TS4500 & LTO tape



Background to our latest cost model



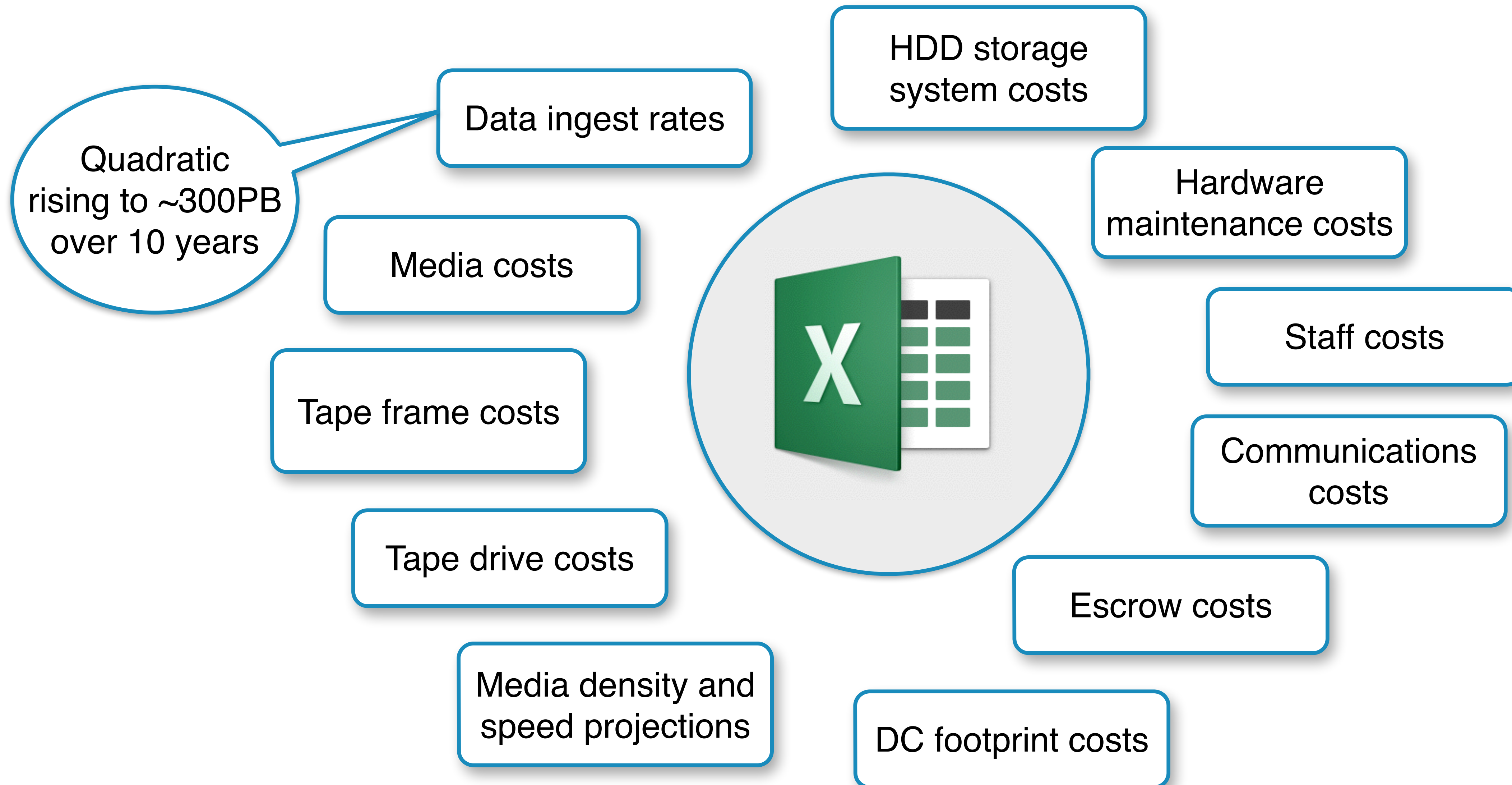


Questions...

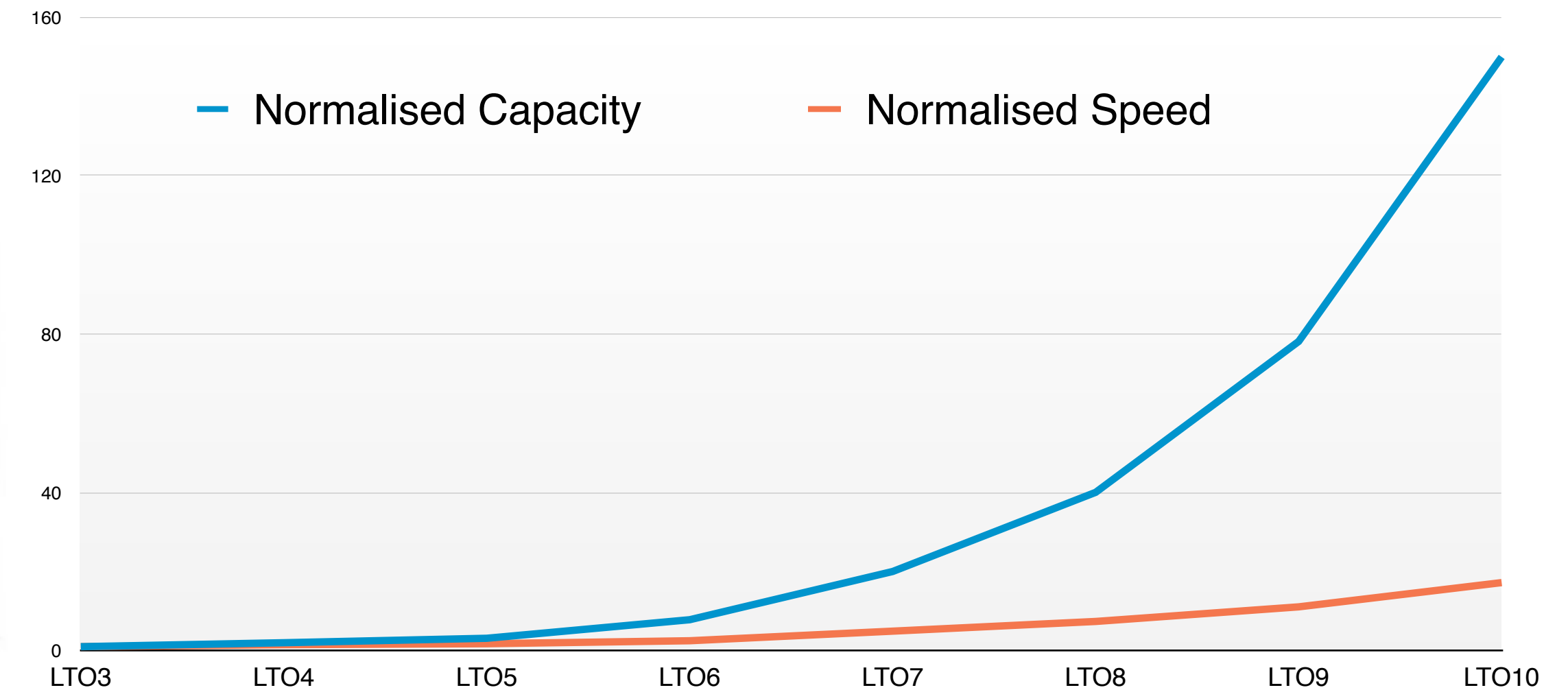
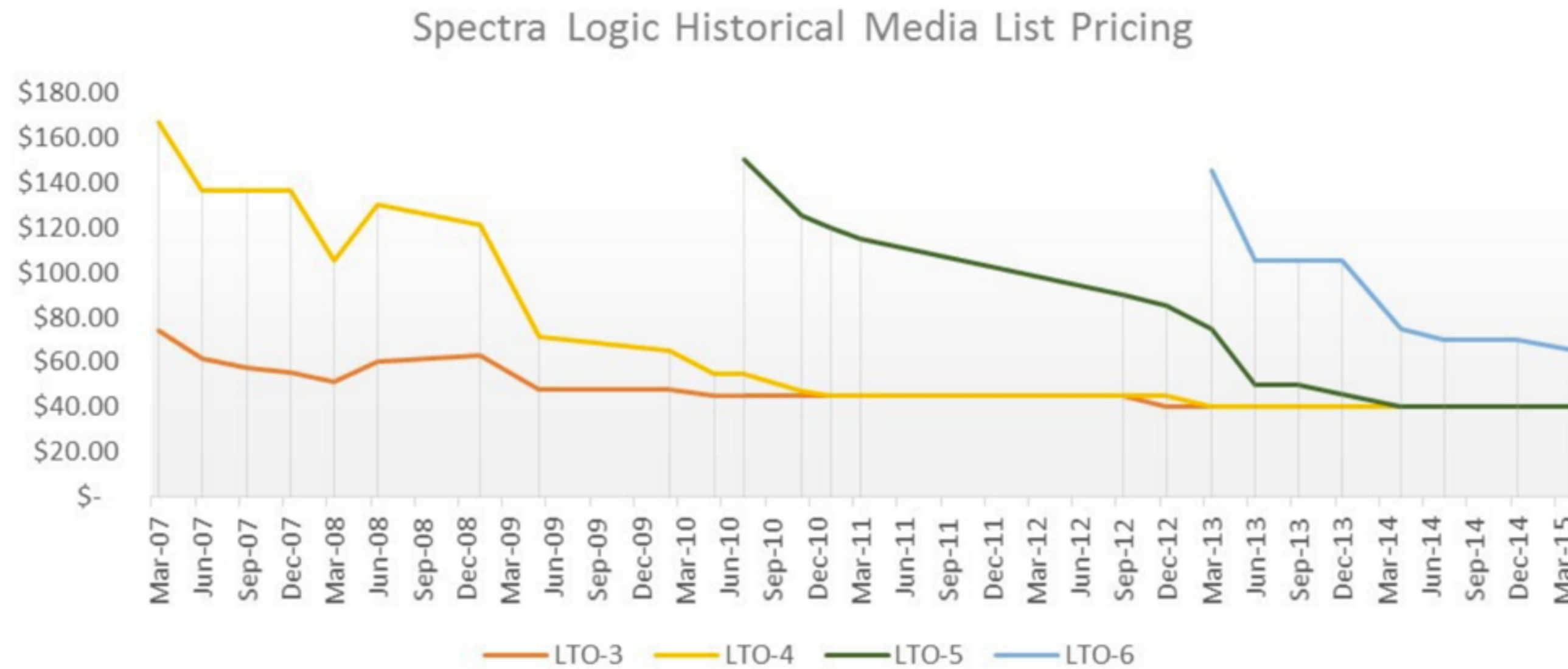
- Nearline tape or HDDs for the DC copy of data?
- Enterprise tape or LTO tape?
- When to migrate between tape generations?
- Use of commercial CoLo DCs or building our own facility?
- What makes sense for small or large scale archives?
- Should we use third-party cloud services?
- ***How much does it cost to preserve a TB of data for the next 10 years?***



Key drivers in the model



LTO tape generations



<https://edge.spectralogic.com/index.cfm?fuseaction=home.displayFile&DocID=4732>

LTO roadmap <http://www.lto.org/technology/what-is-lto-technology/>

Note that Spectra list prices are higher than high-street media prices due to the Spectra media assurance programme.

Drive prices tend to increase by about 10% each generation

Capacity grows faster than read/write speed

DC cost savings through smaller footprint are offset by the increased price per cartridge in the early days of release - wait until \$/TB parity before adopting new generation

Enterprise media vs LTO



1. Higher capacity media
2. Faster read/write performance
3. Lower DC costs
4. Higher reliability
5. Drive migrations allow greater capacity from older media

Enterprise tape offers no significant cost saving over LTO with higher media and drive costs offsetting savings in DC space.

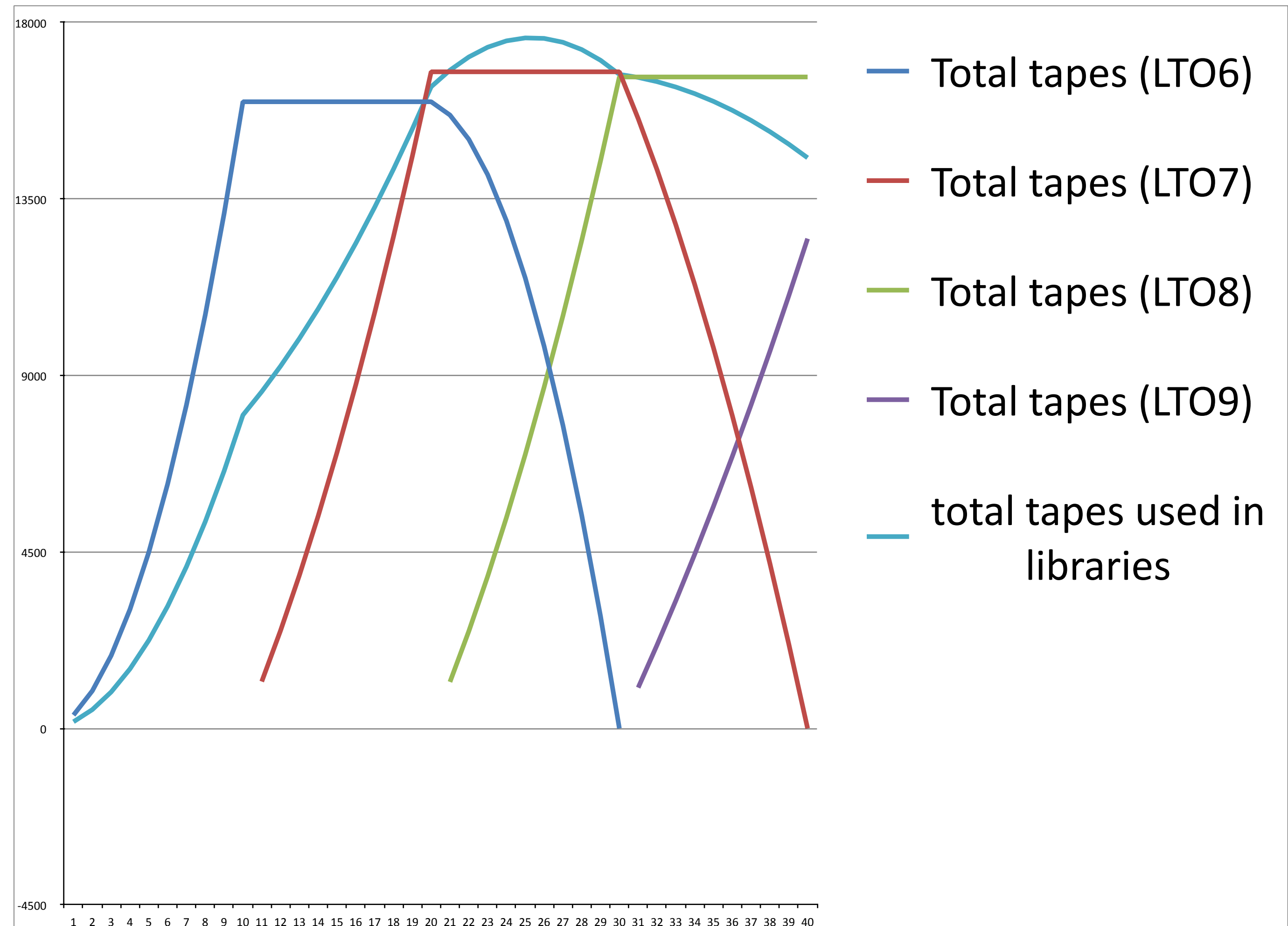


1. More expensive media cost per TB
2. More expensive drives
3. More expensive support and maintenance
4. More drive hours to migrate to make us of 5
5. Difficult or impossible to mix LTO and Enterprise within same library

Media migration strategies

The optimum strategy has three components:

1. Use all LTO generations in the library and ingest new data onto the latest generation.
2. Migrate data every other generation, i.e. LTO5 to LTO7, LTO6 to LTO8.
3. Migrate data slowly rather than ASAP, i.e. if a generation is in play for say 2.5 years when ingesting data, then take 2.5 years to migrate it off again.



Media Migration

Total
Data

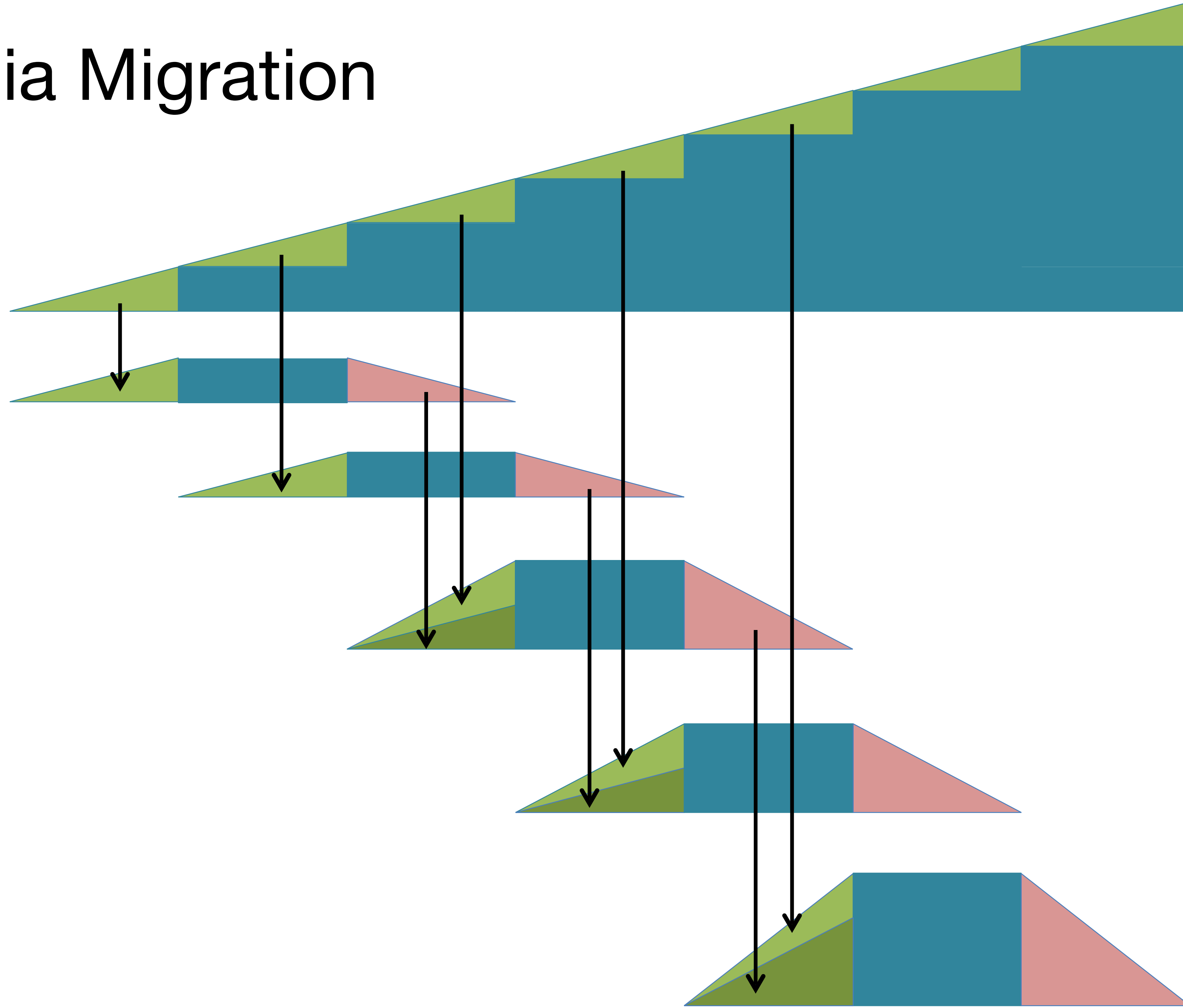
LTO5

LTO6

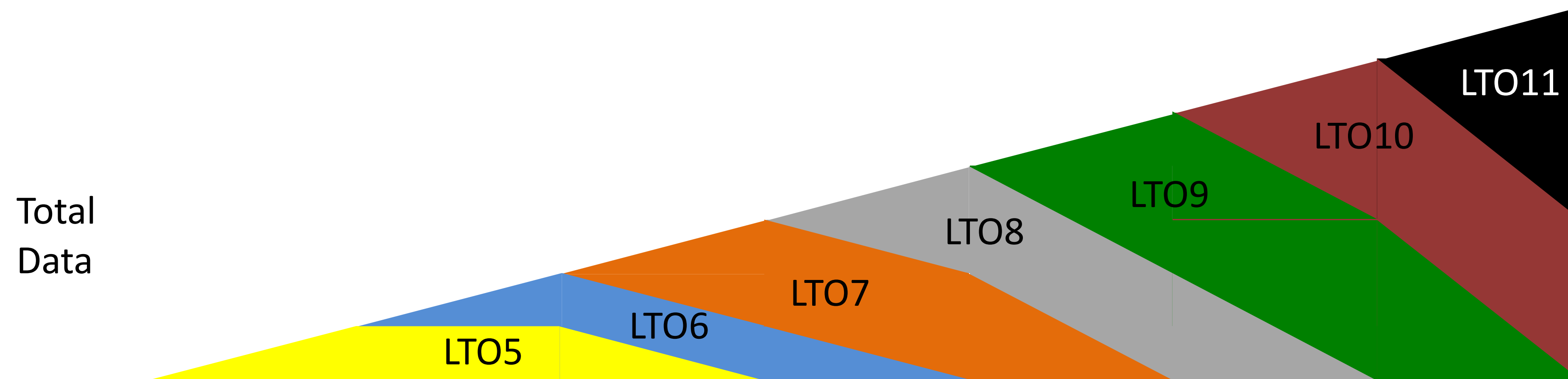
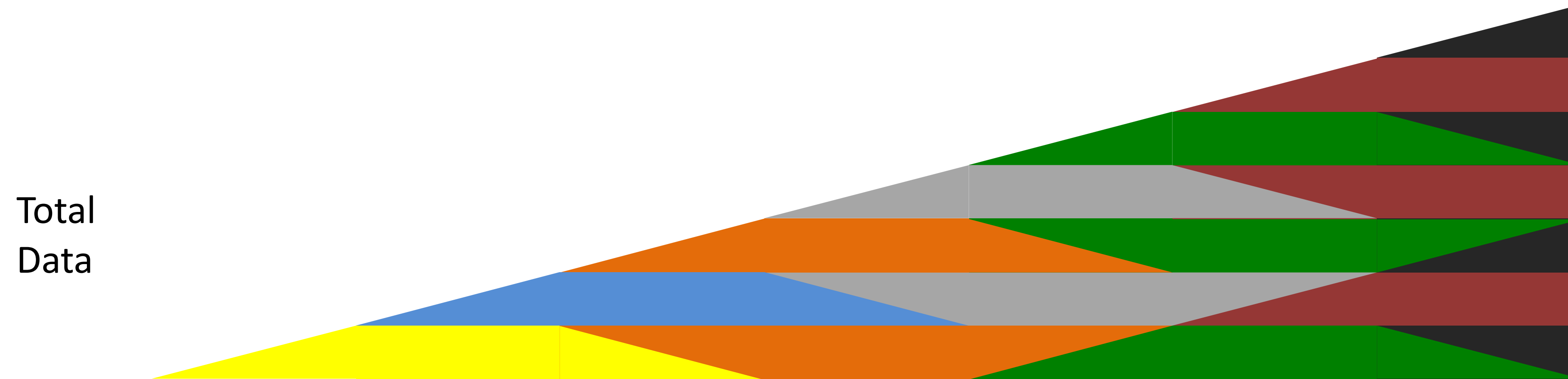
LTO7

LTO8

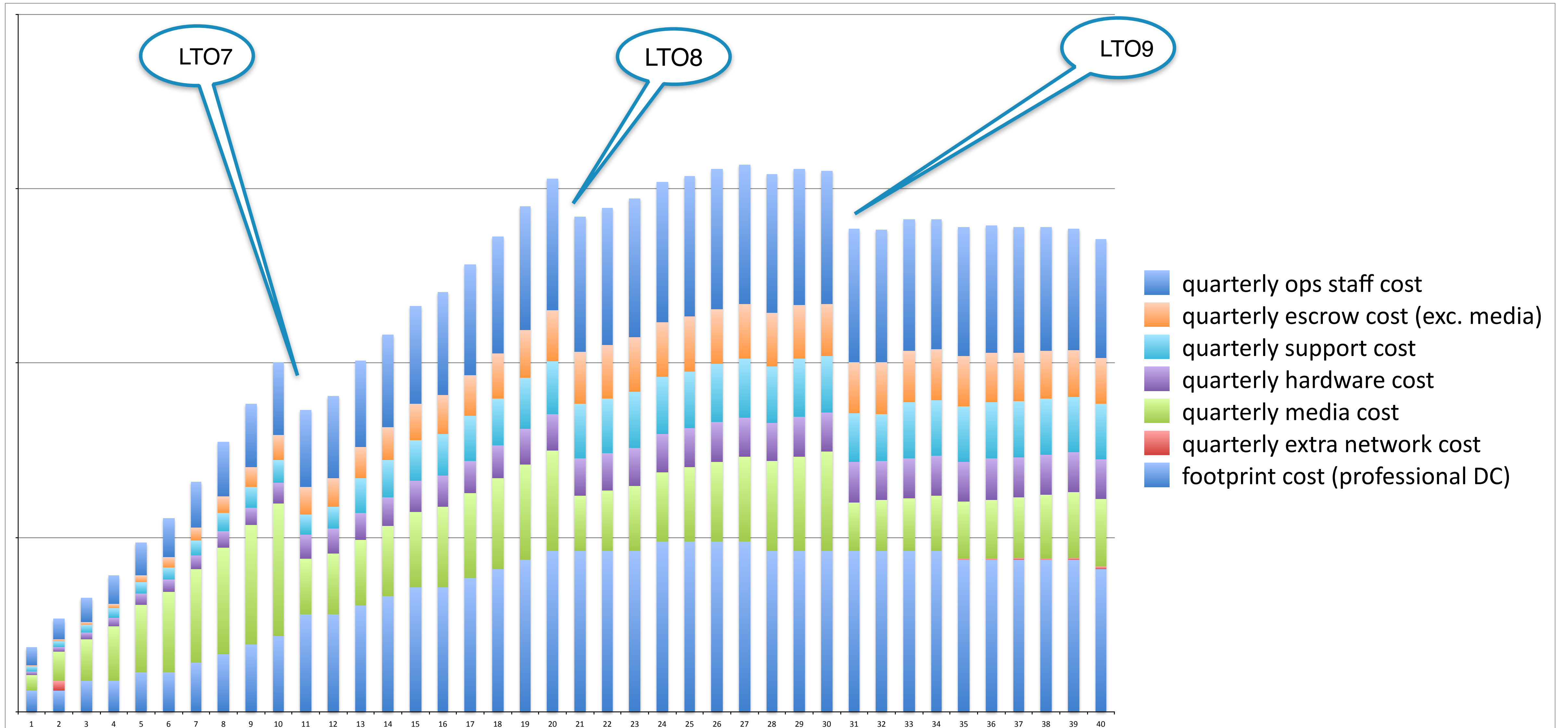
LTO9



Media Migration



Overall costs - DC CoLo



Conclusions



Conclusions

Based on our costs and data growth rates

- LTO is more cost effective than Enterprise
- DC footprint costs are highest proportion, followed by media
- Adopt new LTO generations as soon as \$/TB parity
- Migrate every other generation
- DIY DC costs are dominated by network costs
- 1st year costs fall by 98% by the 10th year
- Finally...



How much to store 1TB for 10 years?

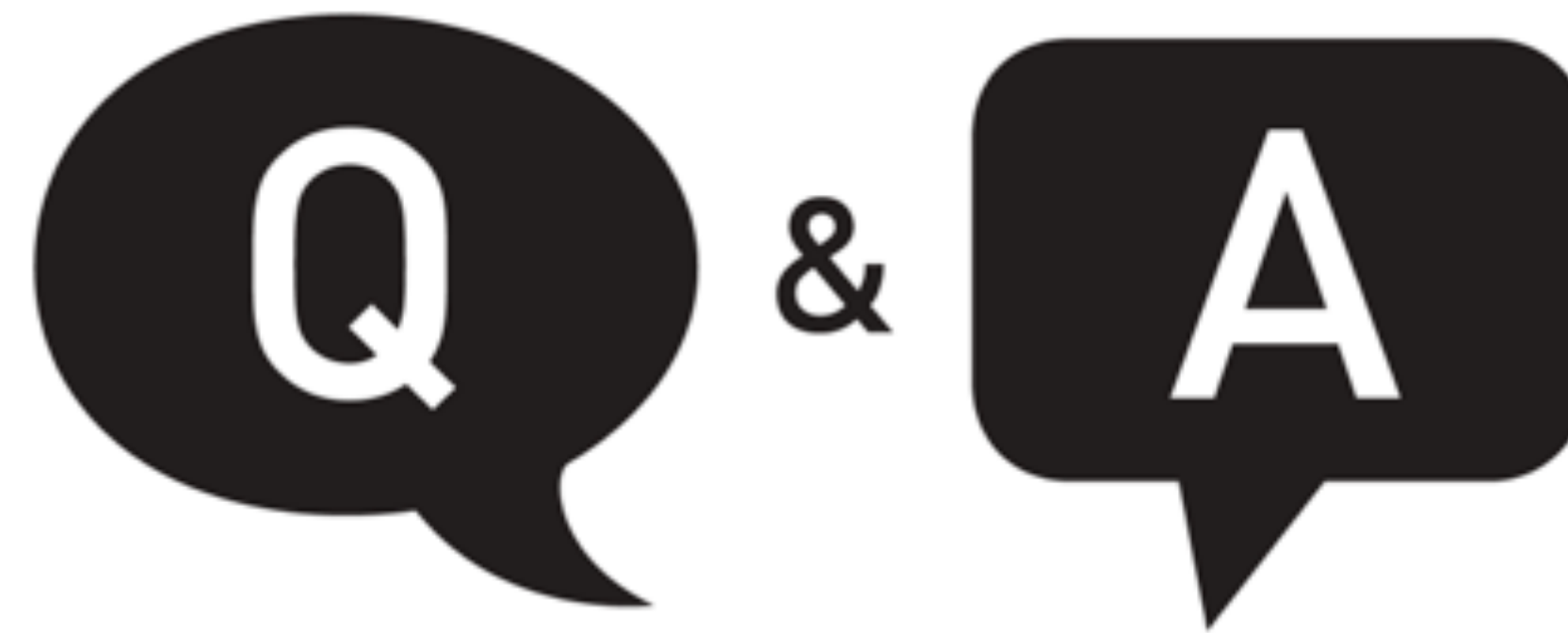
It's still hard to tell...

Assuming "scale"...

Less than \$600...

Probably!





jim.cook@arkivum.com

+44 7785 558432

+44 1249 405060

www.arkivum.com

