ARCHIVAL AND BACKUP IN THE CLOUD

Jason Adrian
Principal Hardware Architect
Azure Storage

Microsoft Azure

The cloud is different….

- On-prem expectations and limitations do not always hold true in the cloud
- Cloud storage offerings break the mold of what is possible with data storage
- We offer a native tiered storage model with consistent APIs across all tiers
- Customer expectations of the cloud are no longer tied to a type of hardware, or a specific software package

Ed Childers from IBM said there are no data elves in the cloud…

I disagree, many of them wear Azure shirts

How do I know?

On December 13, 2017 Azure launched the lowest cost archival solution in public cloud, at **half** of our nearest competitors price

Premium

Low and consistent latency data

Hot

Frequently accessed data

Cool

Less frequently accessed data

Archive

Rarely accessed data

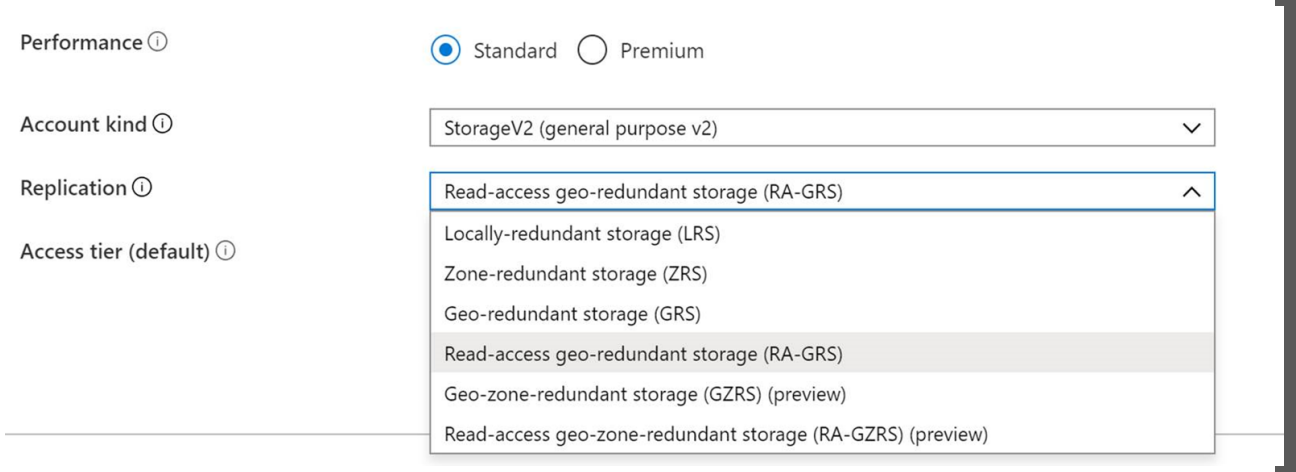
- Our goal is to design an infrastructure and service offering to enable you to focus on your core business, not the business of managing complex hardware and software
- We hide the complexity and offer simple APIs to manage your data
- We ensure the durability of your data, ensure media health, and avoid any concept of technology migration

- But the cloud isn't magic…
- We use many of the same hardware components that are used on-prem. This includes all the tradeoffs in performance and cost, as well as performance and latency expectations.
- We use the same or similar SSDs, HDDs, and tapes that can be purchased on the open market
- We do however design ASICs or utilize FPGAs where we see the ability to be even more efficient or performant than commodity hardware would allow

- We have some benefits unique to cloud scale:
  - Design our own hardware
  - Design amazing software
    - due to the volume of data, we have to invest in differentiated software capabilities
    - utilize efficient erasure codes to enable durability far greater than 2 or 3 copies
  - Influence the industry due to our volume
  - Leverage large infrastructure capabilities to offer performance that is nearly impossible to match on-premises (not many companies have millions of cores of compute, or millions of spindles of HDDs, for example)

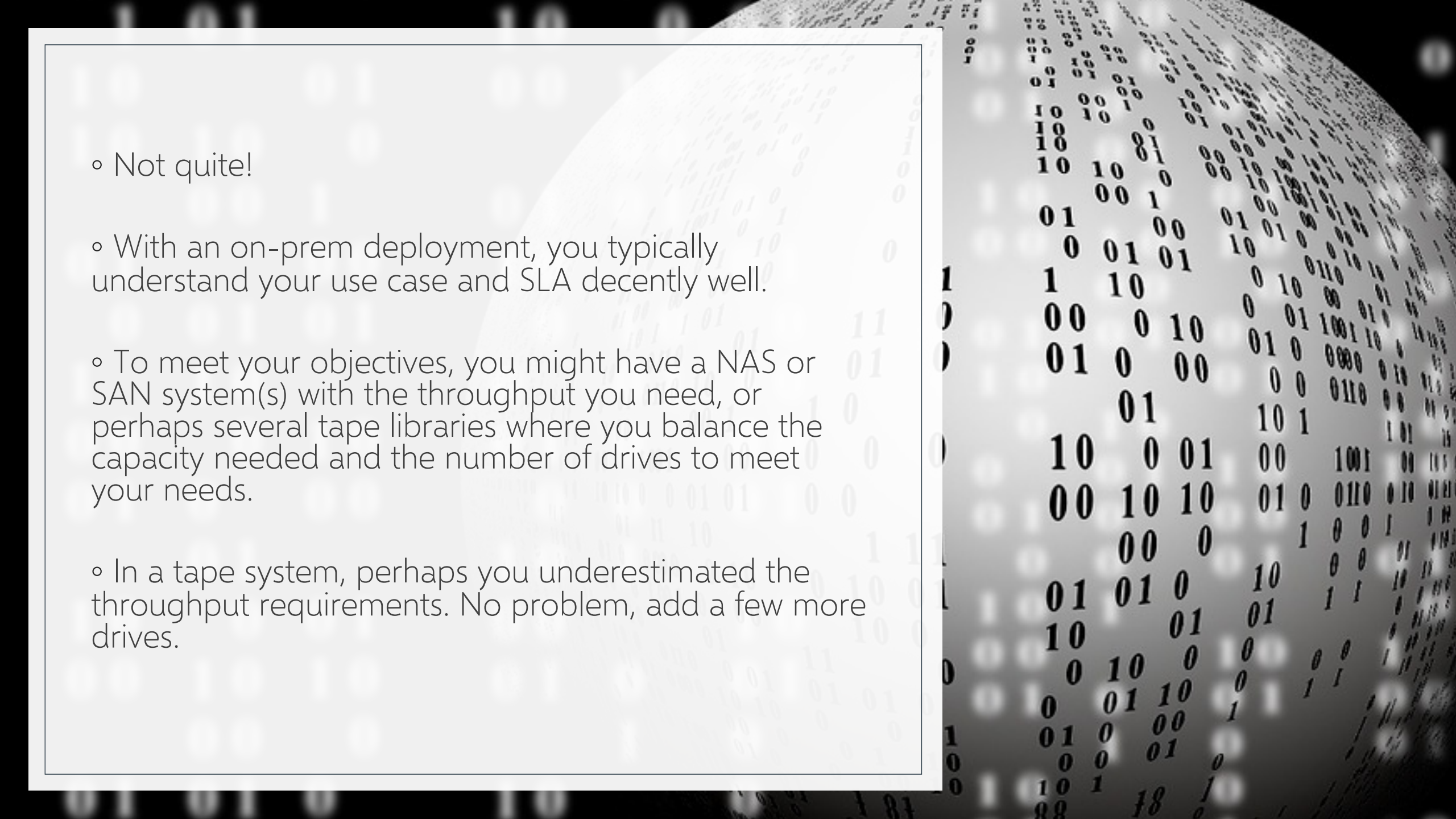
We can offer redundancy options, with the click of a button!

Before cloud, you needed multiple physical locations, multiple hardware deployments, and a large team to make this work. Not very flexible, and certainly not a quick deployment strategy. Storage companies realized this issue, and started selling software to manage data migration between dis-similar storage systems

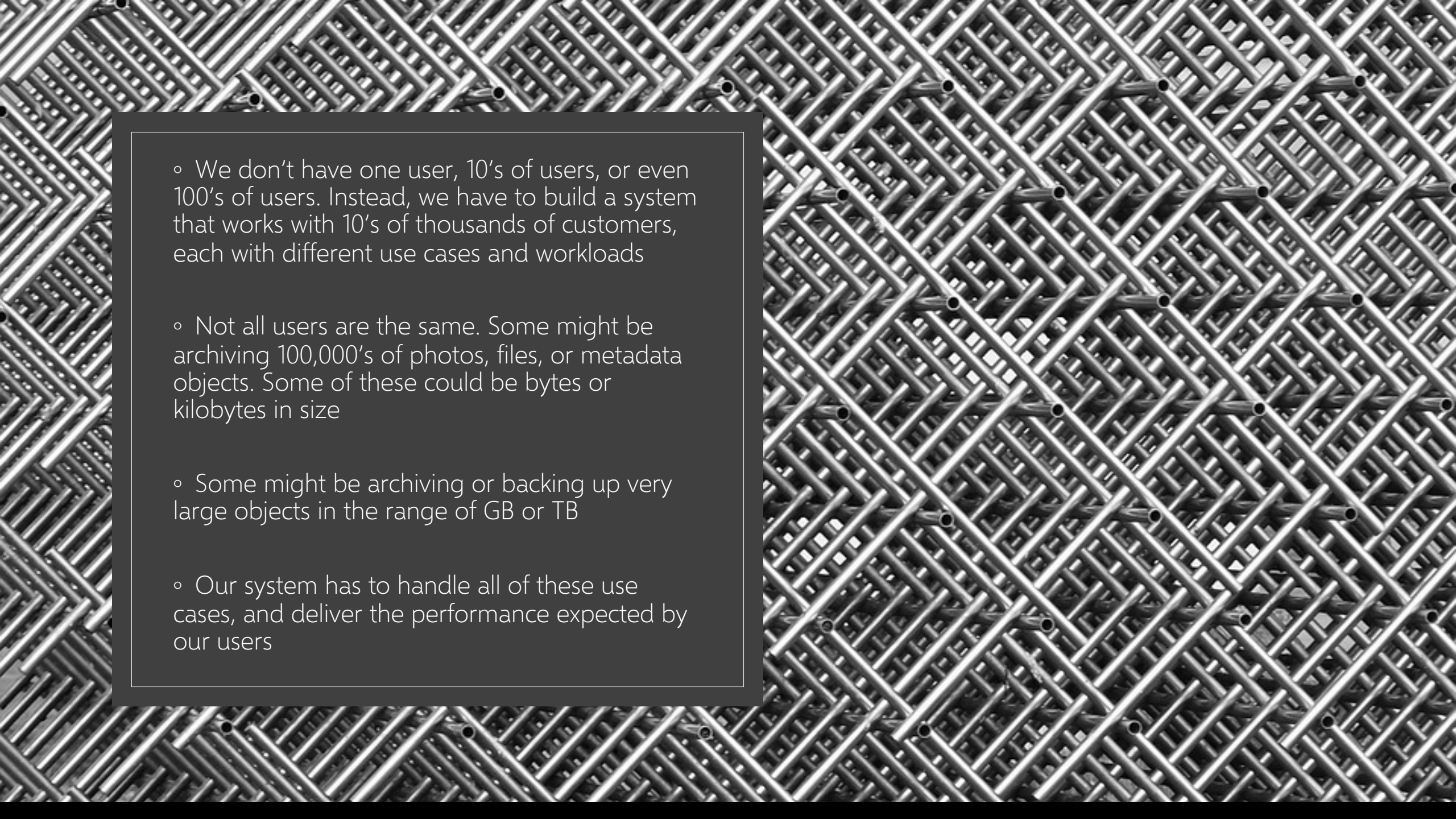


- Cloud storage is often called out as too expensive, or you see TCO comparison charts showing on-prem being a lot cheaper
- But are you comparing apples to apples?
- In Azure Archival Storage, your data is durable and available, with multiple copies and/or erasure coding
- What is the cost of a migration? Infrastructure downtime?

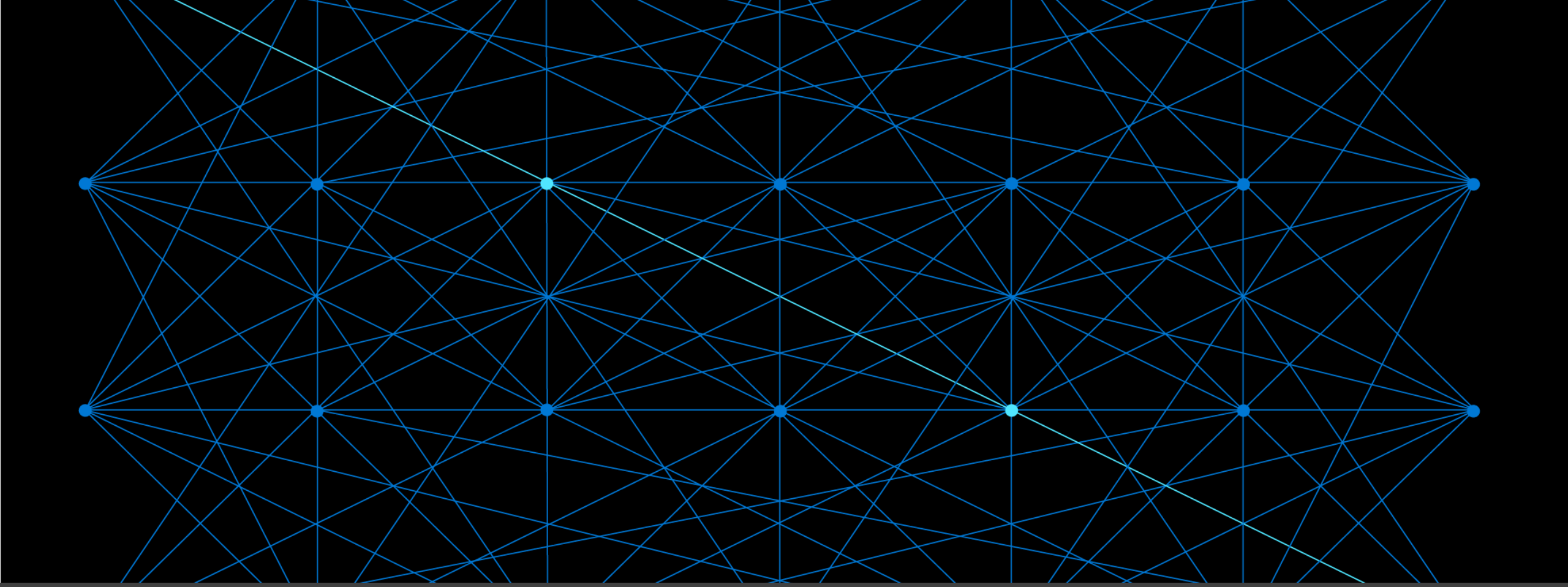
SO DESIGNING CLOUD STORAGE IS SIMPLE, RIGHT?



- Not quite!
- With an on-prem deployment, you typically understand your use case and SLA decently well.
- To meet your objectives, you might have a NAS or SAN system(s) with the throughput you need, or perhaps several tape libraries where you balance the capacity needed and the number of drives to meet your needs.
- In a tape system, perhaps you underestimated the throughput requirements. No problem, add a few more drives.



- We don't have one user, 10's of users, or even 100's of users. Instead, we have to build a system that works with 10's of thousands of customers, each with different use cases and workloads
- Not all users are the same. Some might be archiving 100,000's of photos, files, or metadata objects. Some of these could be bytes or kilobytes in size
- Some might be archiving or backing up very large objects in the range of GB or TB
- Our system has to handle all of these use cases, and deliver the performance expected by our users



WE SOLVE THE PROBLEM WITH SCALE AND CREATIVITY

- Let's talk tape….
- Tape media density is increasing, which is great!
- So that translates to a big savings, right?
- It's not that simple….

- We focus on the TCO of technologies and the infrastructure built upon it
- When the capacity of a cartridge doubles, we don't realize that density savings in full
- With the higher capacity comes an increased need for more drives to balance the throughput and IOPS needs
- The tape ecosystem often makes small improvements generation to generation, but every once in a while we get a doubling of channels and thus a large boost in raw throughput
- The seeks, however, don't get better
- Tape is great for large block IO, but if you need multiple smaller objects on tape, you can easily spend 10x the time seeking vs reading or writing
- This is where we optimize our infrastructure to offer a more performant archival service than a dedicated tape deployment alone might provide

- Tape as a media has a solid roadmap to higher capacity due to the sheer amount of film inside of a cartridge
- A kilometer of tape (give or take, LTO vs Enterprise) has orders of magnitude more surface area than an HDD, so there is incredible potential
- The HDD industry has consistently demonstrated their ability to innovate and scale capacity, but loses the edge in $/TB to tape
- HDDs are on the cusp of energy assisted recording to drive the next growth curve
- Tape has an opportunity to get ahead of HDD in per-unit capacity and TCO
- HDD capacities of 20TB have already been announced! Tape needs to keep up!

Tape Challenges

Tape also carries with it some stringent environmental specifications that are problematic for adiabatic data centers, but fairly common for traditional data centers

Tape needs to keep its competitive advantage in cost per TB – 2018 and most of 2019 wasn't a good year for the tape ecosystem, but hopefully we are back on track

For LTO, we need to see LTO 9 hit 24TB and maintain this strong growth in capacity with future generations

The cost curve must be **substantially** lower than HDDs due to the offline nature of the technology

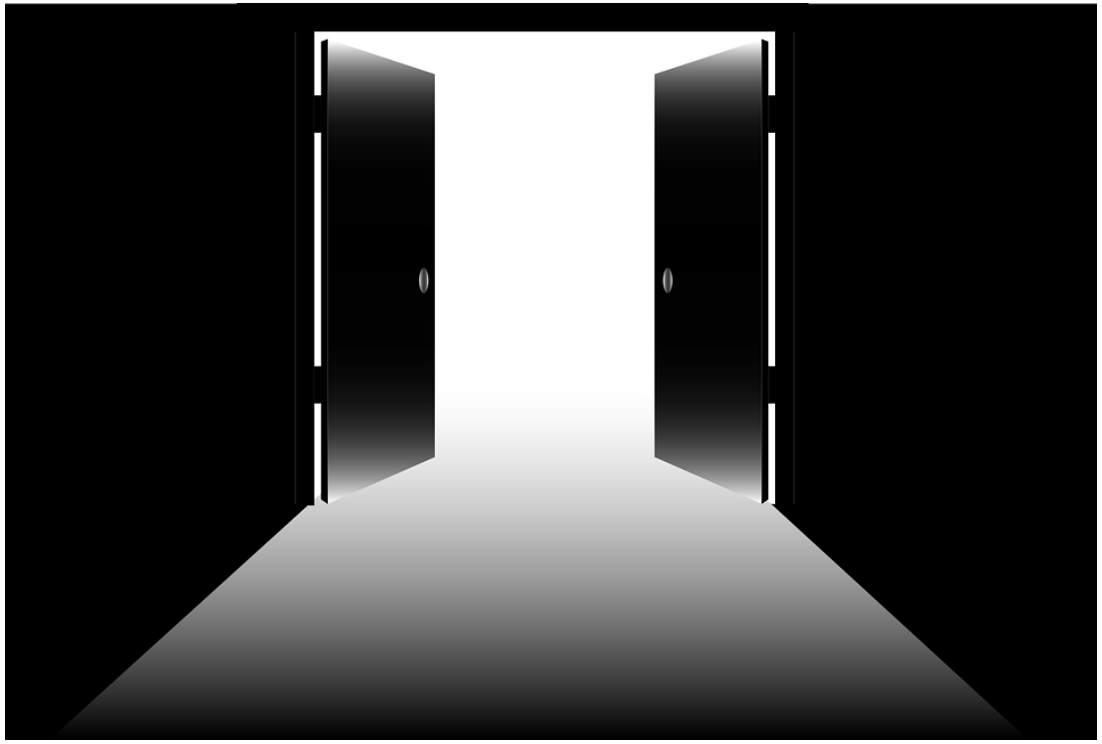
Tape Opportunities

The LTO and Enterprise ecosystems need to drastically push for both density and throughput

We need to see cartridge capacities double whatever the latest HDD capacity is

If we can increase the environmental capabilities of the media to meet adiabatic datacenters, and maintain a 10+ year useful life, this will be a game changer

I'm confident that the drive and media technology creators can do this



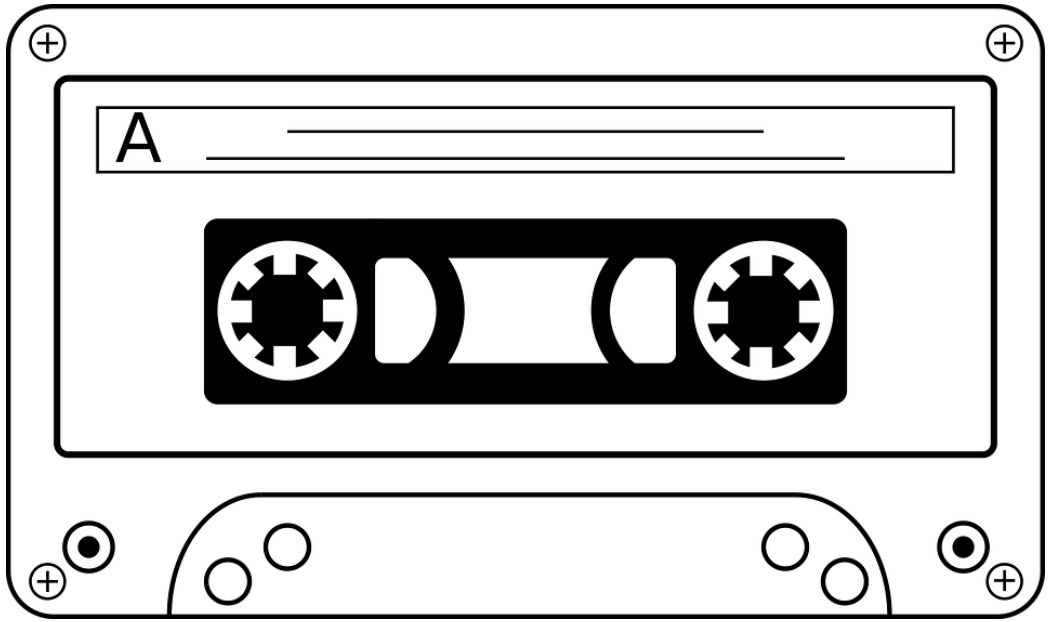
Tape is dead?

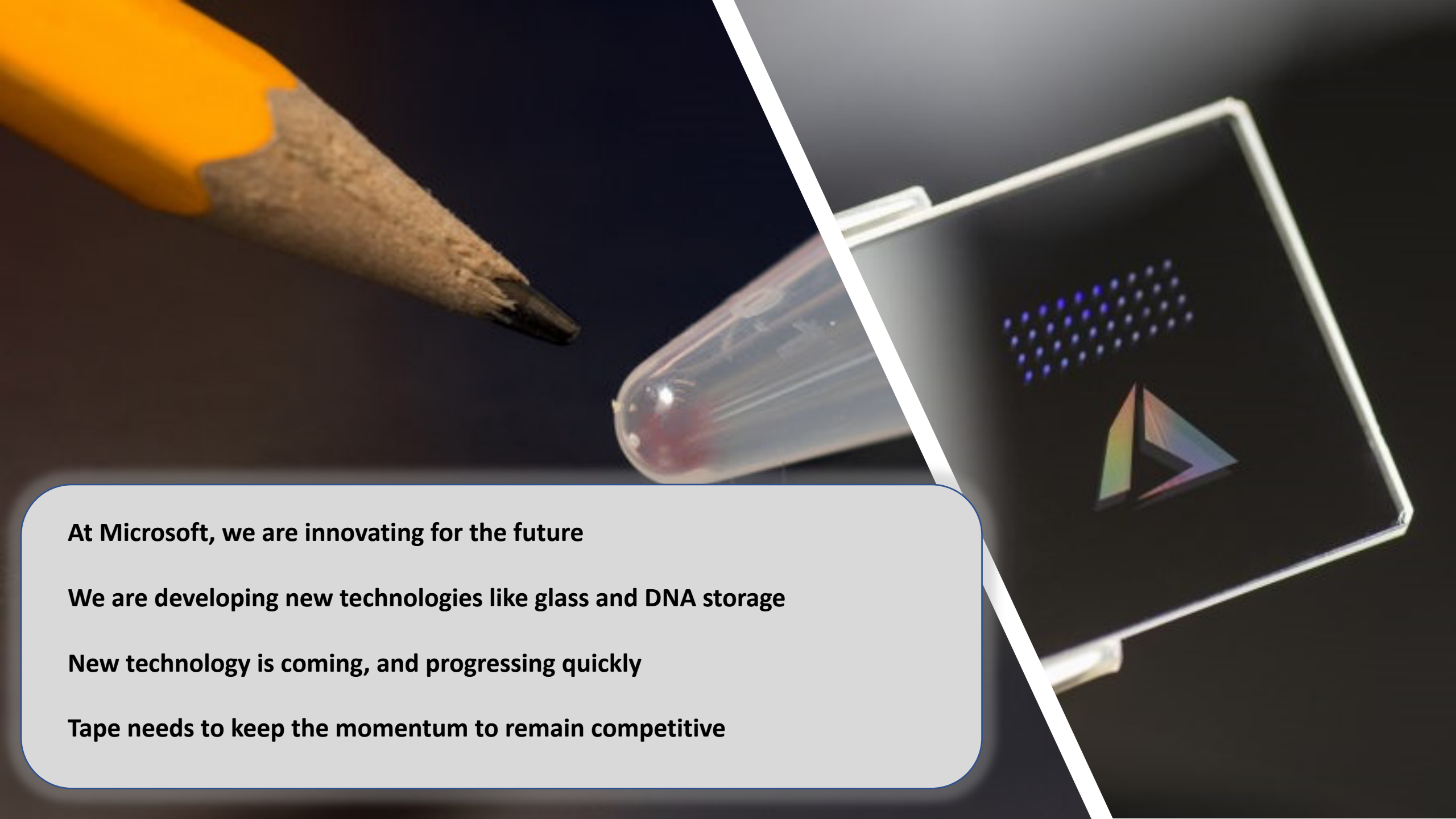
For a technology that people said was dead years ago, tape sure looks alive and well

The tape technology of the 90's and early 2000's is nothing like today's media

If tape continues to deliver on its promise of low cost and high density, it will continue to see more adoption in hyperscale datacenters

For those that think that tape is dead, perhaps it has simply found a new home?





At Microsoft, we are innovating for the future

We are developing new technologies like glass and DNA storage

New technology is coming, and progressing quickly

Tape needs to keep the momentum to remain competitive