The Bit Mountain Research Project

By Shane Hathaway
and the Touchstone team

Story

❖Salt Lake Community College
  – Program Innovation
❖The data must outlive the media

What's a petabyte?

❖ 1,000,000,000,000,000 bytes (quadrillion)

❖ 1,000,000 GB

❖ 1,000 TB

❖ We need to store roughly 18 PB--forever

   – 3 million films

   – 1000 images per film

   – 6 MB per image

❖ Will store even more over time

❖ This **is** the backup

Methods Considered

- DVD
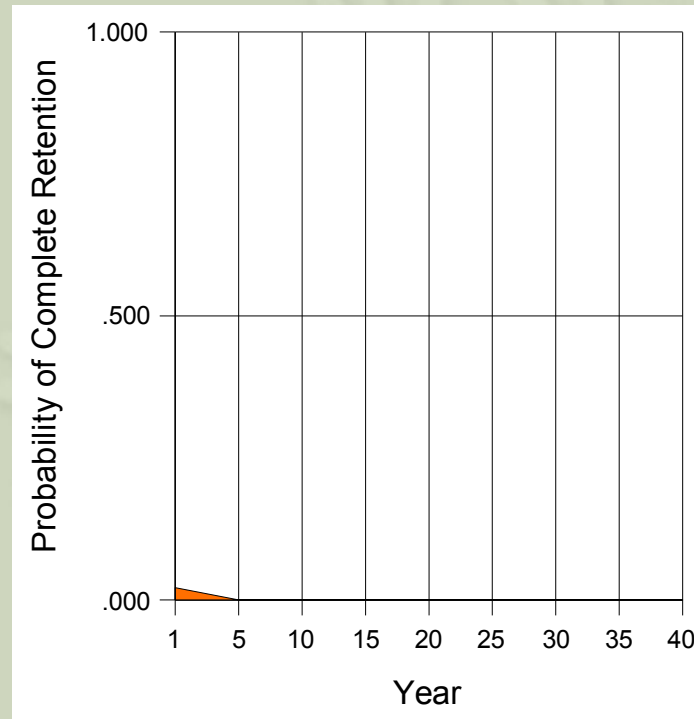- Digital Tape
- UDO
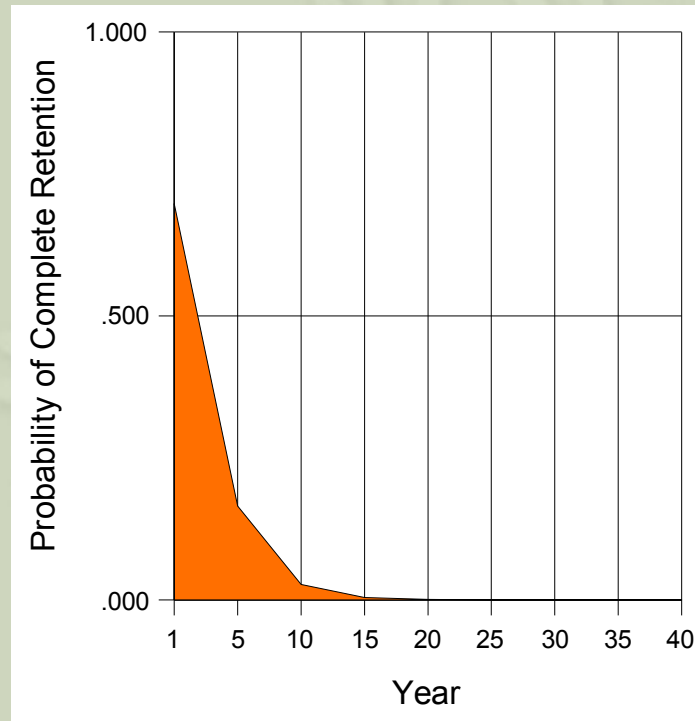- Digital Microfilm
- Silicon Etching
- MAID

DVD

❖18 PB x 3 / 4.7 GB = 11,500,000

❖Refresh once per year: 44,000 DVDs per day

❖1 in 100 fail unexpectedly per year
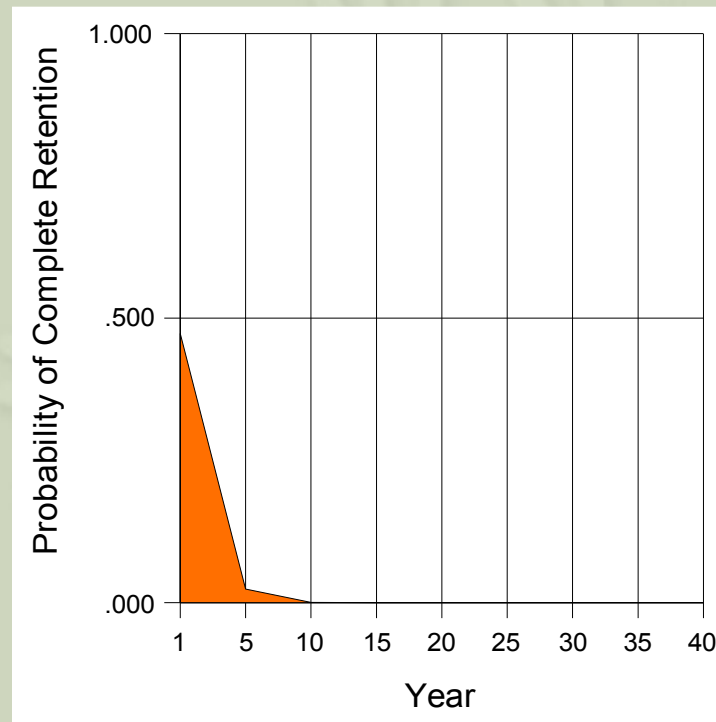
Digital Tape

❖ 18 PB x 3 / 400 GB = 135,000

❖ Refresh once per year: 520 tapes per day

❖ 1 in 50 fail unexpectedly per year

UDO

❖ 18 PB x 2 / 30 GB = 1,200,000

❖ Refresh every 5 years: 923 per day

❖ 1 in 2000 fail unexpectedly per year

Digital Microfilm and Silicon Etching

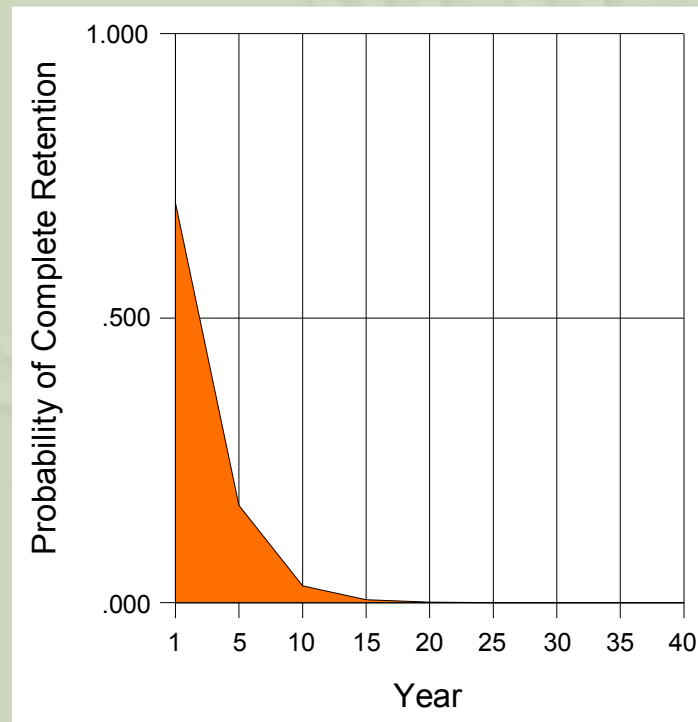❖ We don't have the equipment to test these yet

Replicated MAID

❖18 PB x 2 / 400 GB = 90,000

❖Refresh every week

❖1 in 50 fail unexpectedly per year

Distributed File Systems

❖ Considered:
- MogileFS, Coda, Andrew, Lustre, Global, Google, Oracle Cluster, Ibrix

❖ These are oriented for speed before reliability
- They don't solve the problem we need to solve
- May be useful for other parts of the system, but not this part

Forward Error Correction

❖RAID implements simple FEC

– RAID 5: safe to lose any single drive

– RAID 6: safe to lose any two drives

❖More advanced FEC yields much higher reliability

– It's safe to lose any *n* media, where *n* is configurable. Higher values of *n* require more media and processing power.
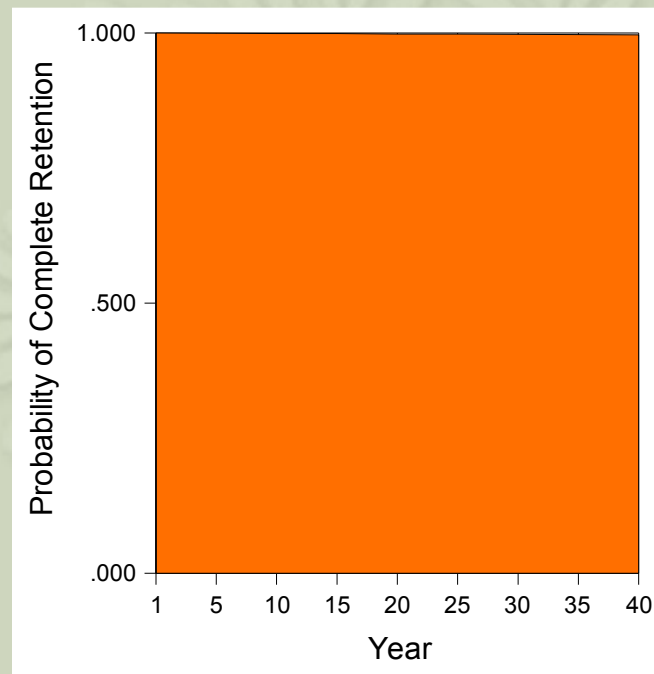
– Chosen algorithm: Reed-Solomon

MAID with Forward Error Correction

❖ 12 data segments, 4 protection segments

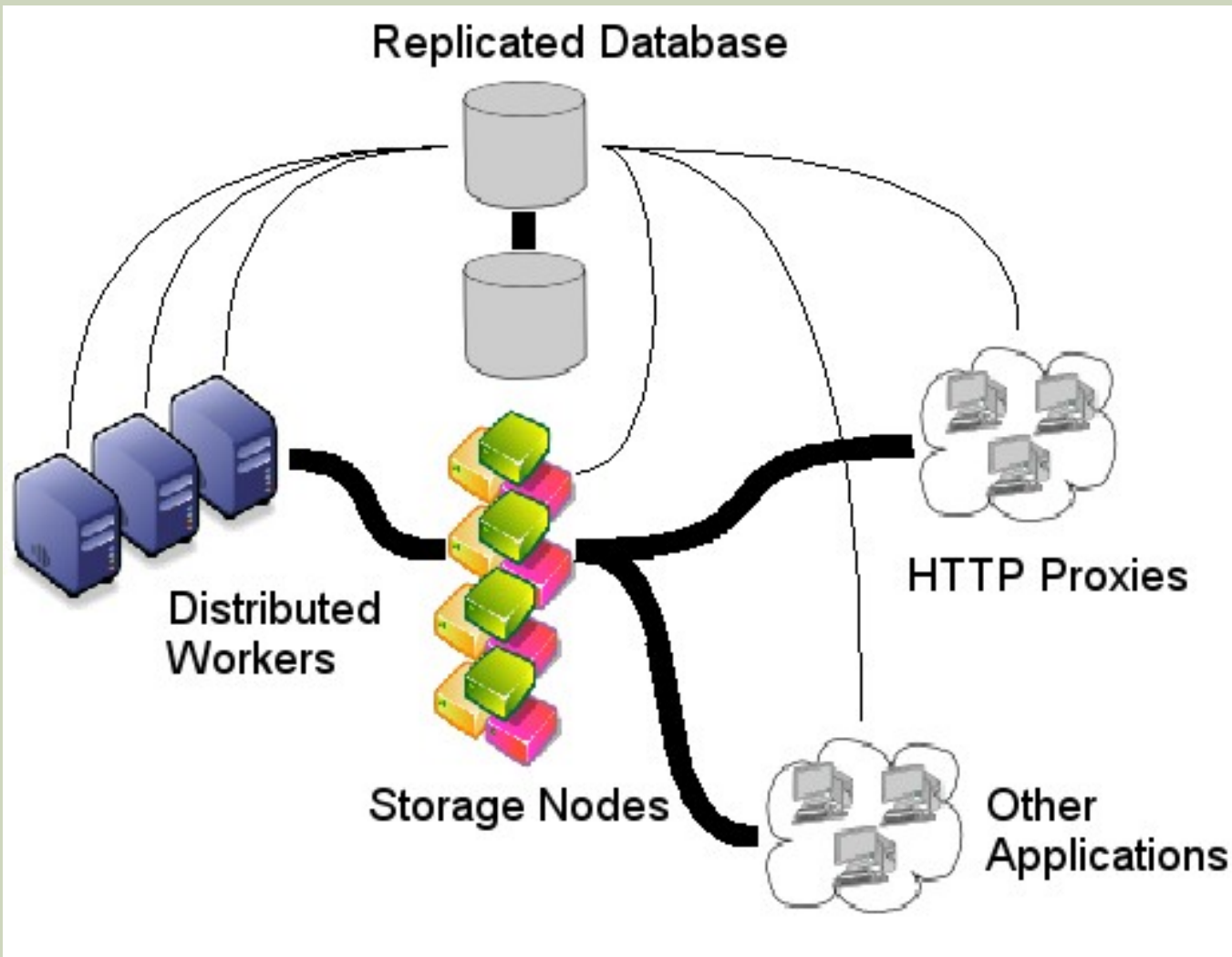❖ 18 PB x 1.33 / 400 GB = 60,000

❖ Refresh every month

❖ 1 in 50 fail unexpectedly per year
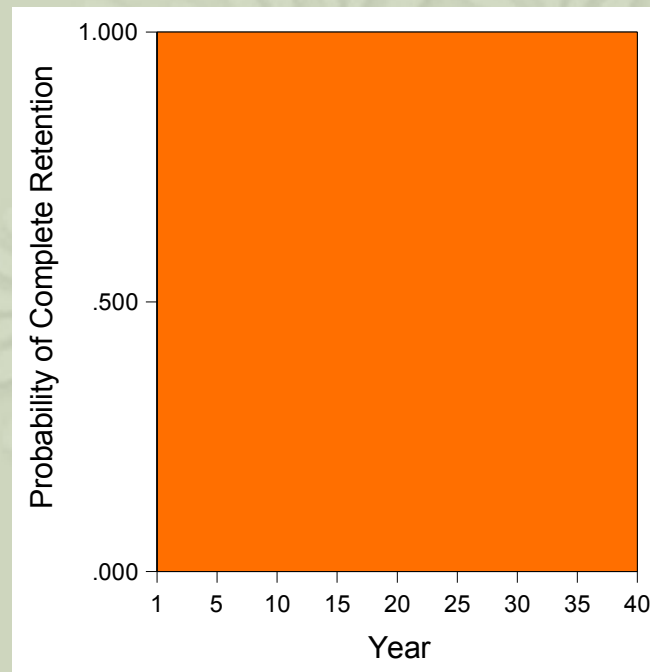
Tapes with Forward Error Correction

❖20 data segments, 7 protection segments

❖18 PB x 1.35 / 400 GB = 60,750

❖Refresh every year

❖1 in 50 fail unexpectedly per year

Bit Mountain Prototype

Bit Mountain Features

❖ Self-healing
  – Devices expire and files are re-created automatically on other devices

❖ Clients can store 100 MB per second
  – Faster than a single hard drive

❖ Distributed and fault tolerant
  – One important exception: the database. But we have ideas on how to fix that.

❖ Open protocols and formats

Future Directions

❖ May fit the Church's needs
  – Or maybe we're learning enough to purchase or build what the Church needs

❖ We hope to release Bit Mountain as open source software
  – We believe it is widely useful
  – Will improve with feedback and more eyes
  – Ideally, we plant seeds now and harvest later

The Mission

"We have seen only the beginning. . . . I am satisfied that this work will go on and touch the lives of millions upon millions of people across the world. And the God of heaven, whose Church this is, will open the way to make all of that possible if you and I and the members of this Church, wherever they may be, will do our part in assisting with that process" (regional conference, Salt Lake City, Utah, May 5, 2002)."