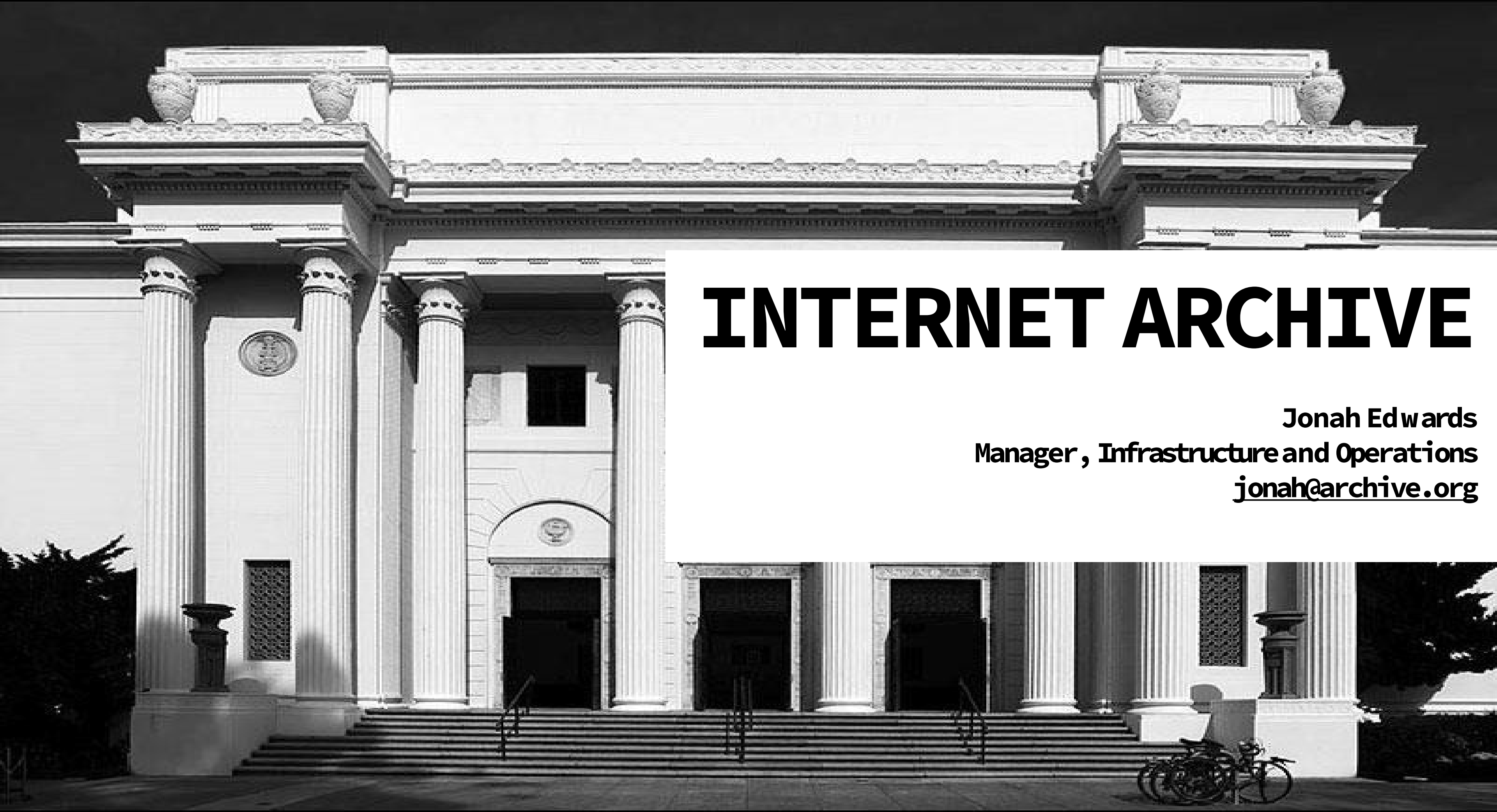


INTERNET ARCHIVE

Jonah Edwards
Manager, Infrastructure and Operations
jonah@archive.org

2022 Materials Update

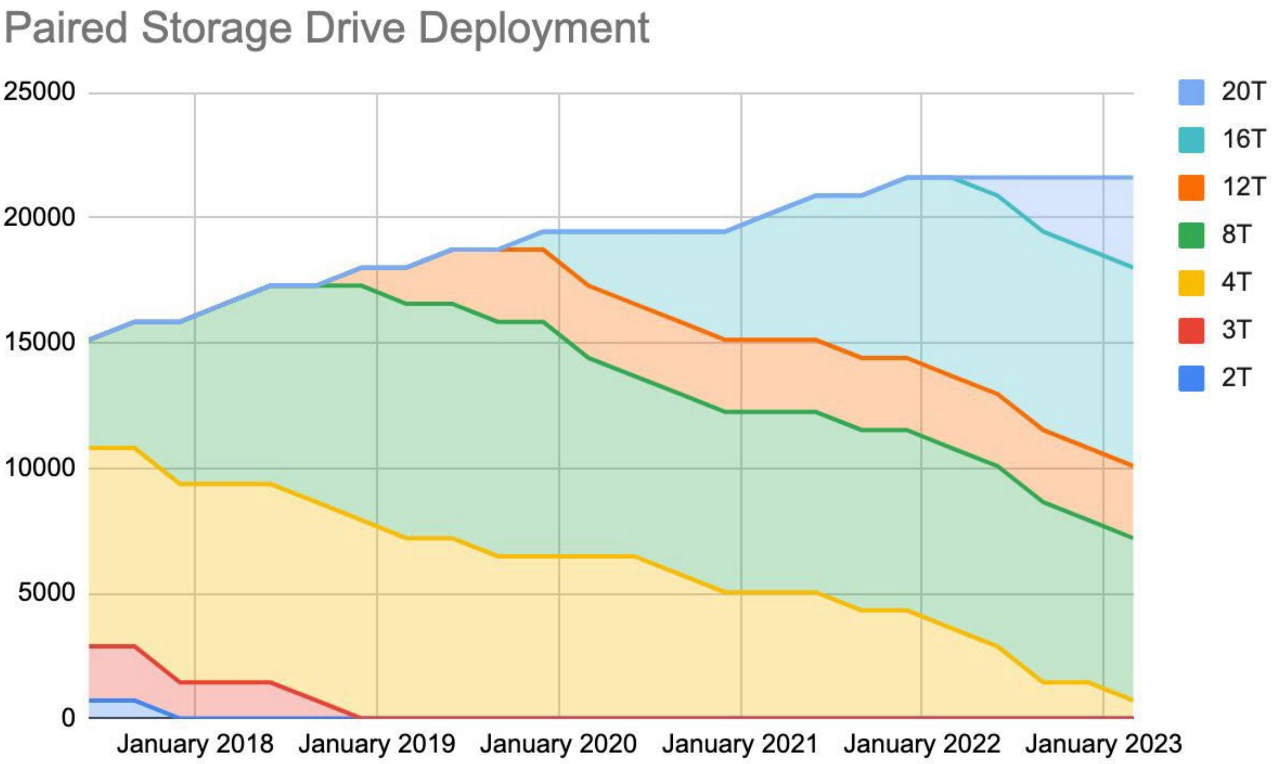
Wayback Machine:
- 735 billion web pages
- 770 million pages captured per day

Collections
- 41 million texts
- 7 million books digitized by us
- 4300 books per day, 18 centers
- 6 million movies (excluding television)
- 2.4 million broadcast news programs
- 14.7 million audio items
- 890,000 software titles, many emulatable
- 4.4 million images
- over 60,000 items in new Amateur Radio collection

2022 Storage Update

Have exhausted current physical footprint and rely on retirement of smaller drive sizes to continue growth

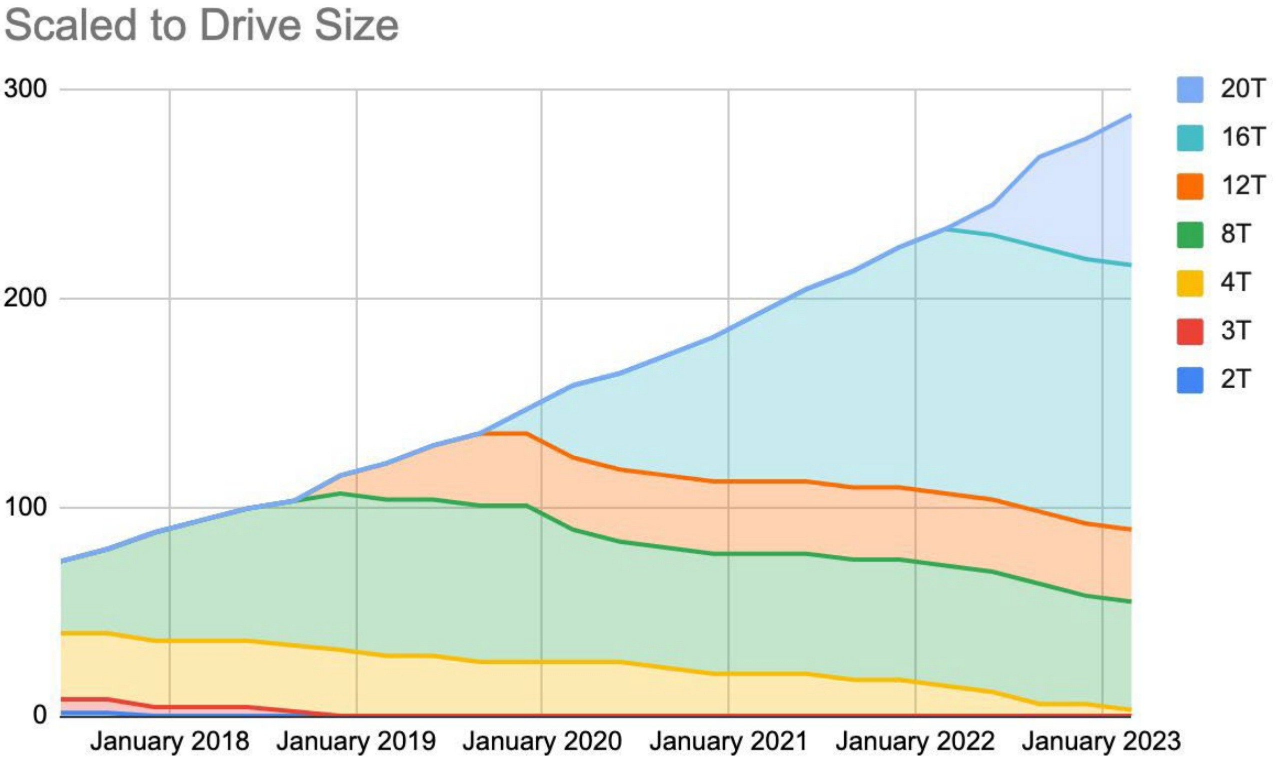
20T deployment over past year has allowed growth of storage platform to keep pace



2022 Storage Growth

Added 23.7 petabytes of material to the stored corpus in 2022, averaging 65 TB/day of data ingest

Total available (not occupied) raw storage volume of ~280 petabytes



IA Canada

- opened new IAC headquarters in Vancouver, BC in June of 2022

- beginning replication of onshore datasets to servers hosted in Canada

- expect to scale this deployment significantly -- have already begun potential deployments with academic partners and at other facilities

- anticipate serving some content from Canada within the year

Next-gen Storage Model

- current storage model: Paired Storage

-- relies on full-disk replication mediated by our internal catalog system

-- advantages include simplicity and transparency

-- replication time for (rare) full-disk failures scales with drive size

- New ZFS-based model

-- server-sized storage pools

-- increased redundancy and ease of maintenance

-- already deployed in Canada -- intent is to build out there while retaining paired storage deployment stateside at this time

Next-Generation Internet Archive

As we continue to scale, moving from a "library that's sometimes closed" stance to a more highly available platform:

- currently averaging 150 Gbps outbound content
- up from 40 Gbps in early 2020
- increasing by ~10 Gbps per quarter

- 18,000 user uploads per day
- 2 million unique website visitors per day
- (whose IP addresses we do not retain)

- anticipate over 500 petabytes outbound in 2023