HED102

Preserving TV News as Cultural Heritage in Perpetuity

Jim Duran
Director, The Vanderbilt Television News Archive, Vanderbilt University

Jon Shaw
University librarian, Vanderbilt University

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

VANDERBILT UNIVERSITY
Jean *and* Alexander Heard Libraries
Vanderbilt Television News Archive

James Pilkington, Administrator 1970 - 1986

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

About the Archive

- Founded in 1968
- A division of the Vanderbilt University Libraries
- A collection of TV news recordings
- AND metadata catalog
- Primary User Groups:
  - Academic Scholarship
  - Film and Media Industry
  - Open to the public

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Demo Video:
Advanced Computing & Society

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Challenges

Storage

Preservation

Automation

Transcripts

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

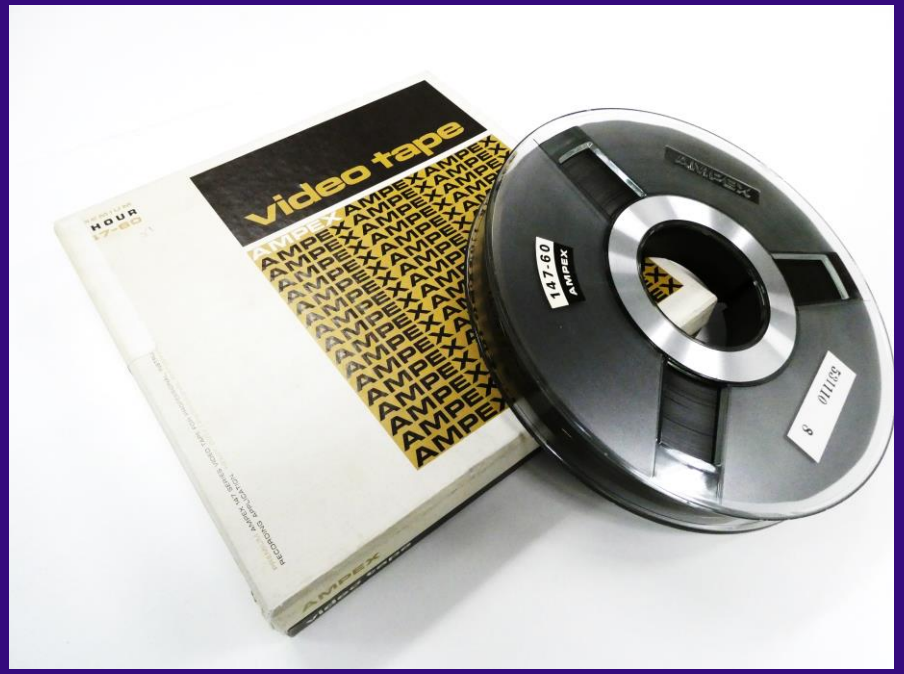
Challenge: Storage

In 2018, The TV News Archive included 175 TB of archival video storage and added 7 TB every year. The department needed sustainable and cost-efficient storage.

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

1st transfer project: 1-inch open reel to 3/4

- AMPEX 1-inch Open Reel



- U-MATIC ¾ Video Cassette

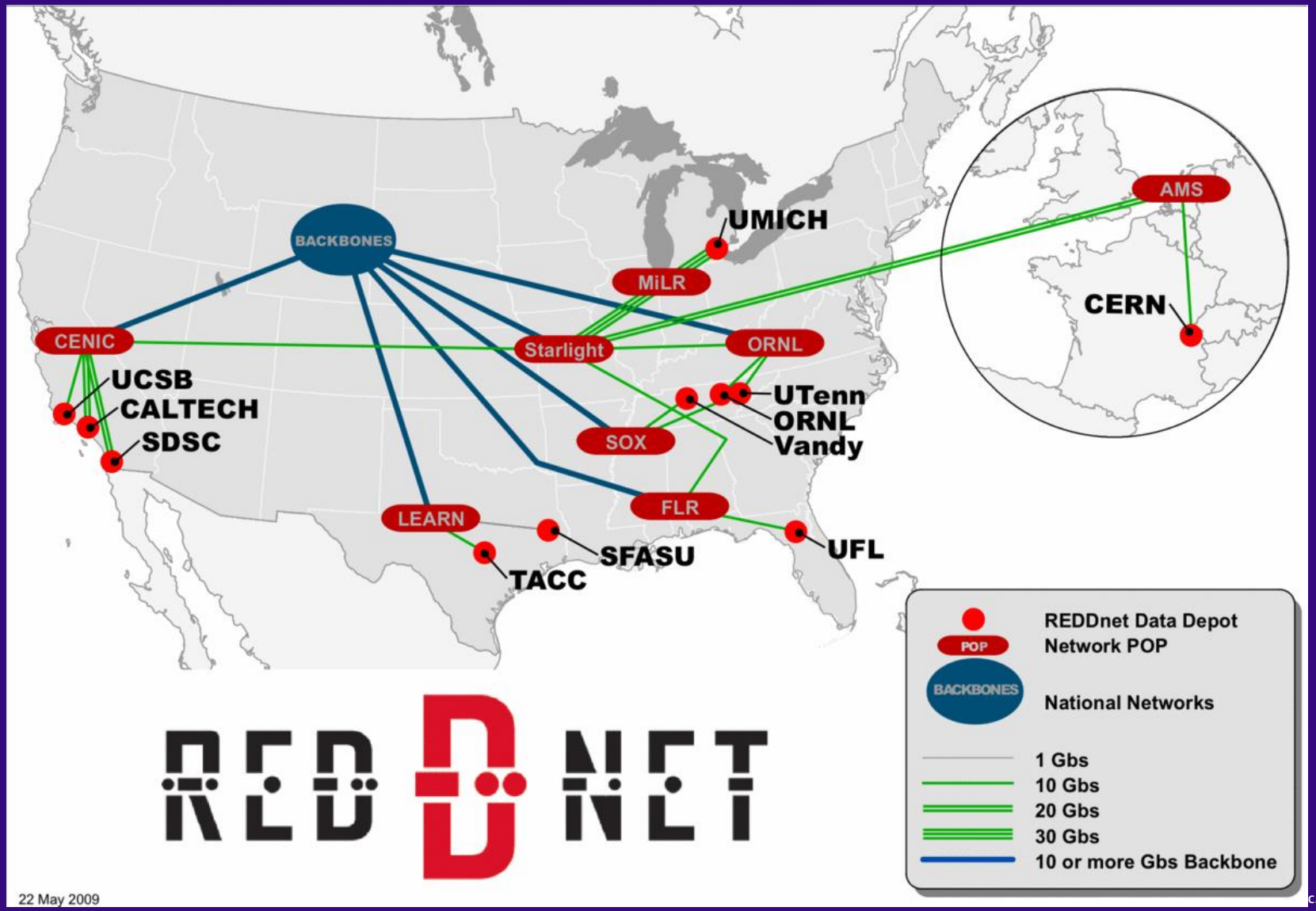
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Second Transfer Project: *Digitization*

- Grant Funded
- 2000 to 2006
- Over 45,000 broadcasts digitized

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Storage Solutions and Requirements

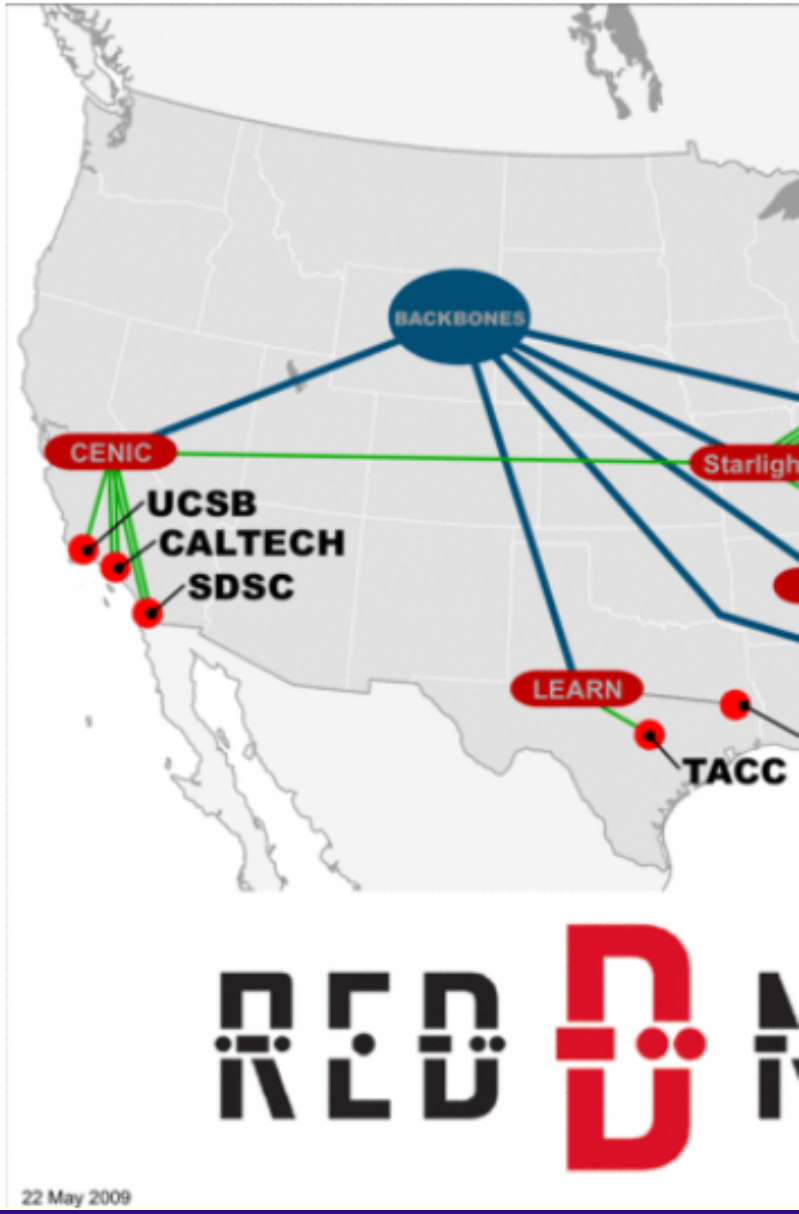


c. or its affiliates. All rights reserved.

REDDnet: Enabling Data Intensive Science in the Wide Area

REDDnet (Research and Education Data Depot network) is an NSF-funded infrastructure project designed to provide a large distributed storage facility for data intensive collaboration among the nation's researchers and educators in a wide variety of application areas. Its mission is to provide "working storage" to help manage the logistics of moving and staging large amounts of data in the wide area network, e.g. among collaborating researchers who are either trying to move data from one collaborator (person or institution) to another or who want share large data sets for limited periods of time (ranging from a few hours to a few months) while they work on it. REDDnet is not designed or intended to be a replacement for reliable archival or long term personal storage and users must make separate arrangements to insure that the data they are sharing via REDDnet's "best effort" storage is also preserved independently with stronger guarantees.

One example comes from the CMS collaboration, a high energy physics experiment that will be taking data soon at the Large Hadron Collider (LHC) at



AWS S3

Started with new files written directly to AWS

Uploaded the 175 TB archive

- Could not use Snowball
- Set to Glacier Deep Archive

Proxy File

- Intelligent Tiering

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Challenge: Preservation

To ensure the long-term preservation of this cultural heritage, VTNA sends a 2nd copy of every news episode to the Library of Congress. The transfer process required an update.

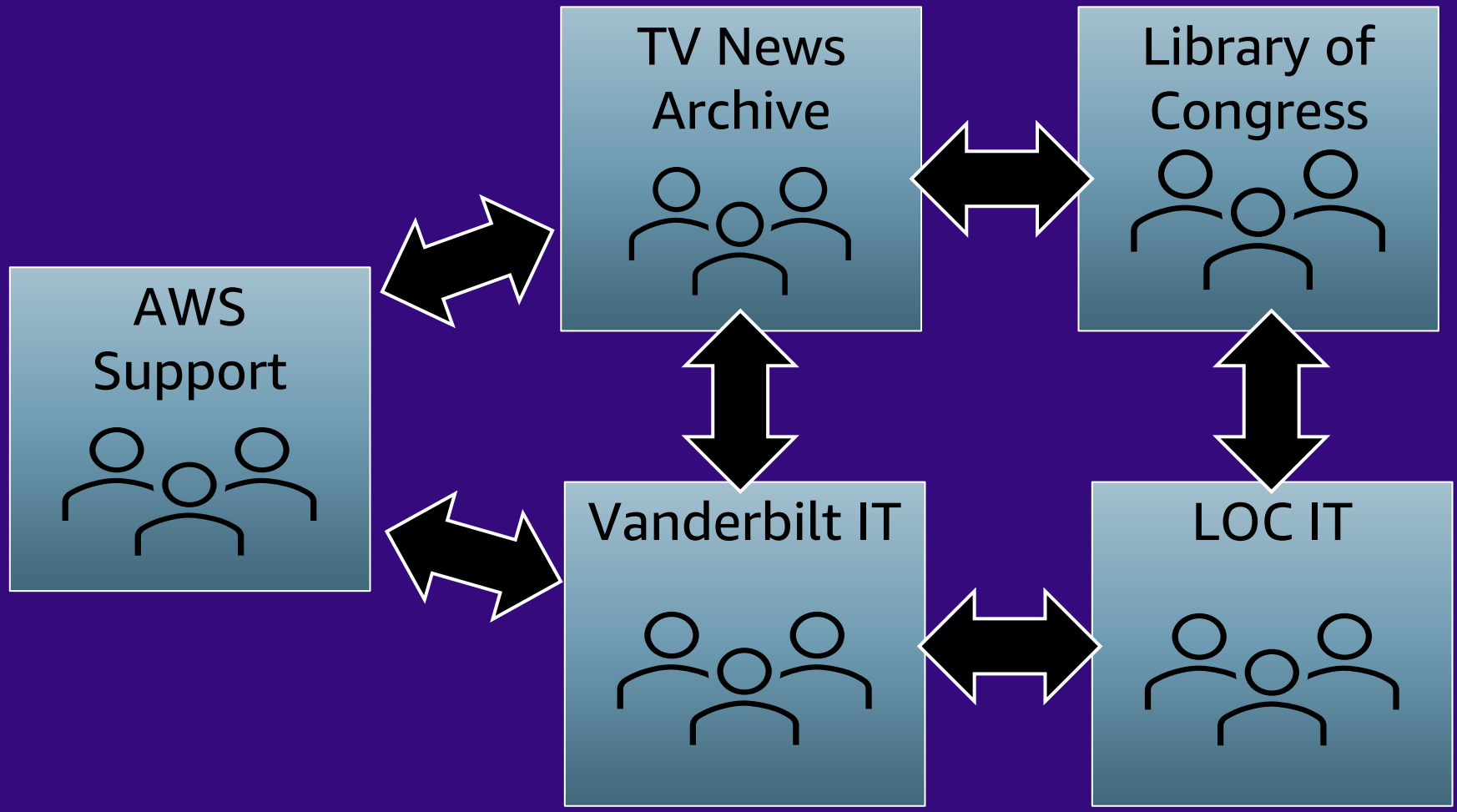
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Direct Transfer to Library of Congress

- Previously used REDDNET
- New system uses bucket to bucket connection
- Verified with checksums
- Initiated with the creation of a new file

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Communication & Teamwork



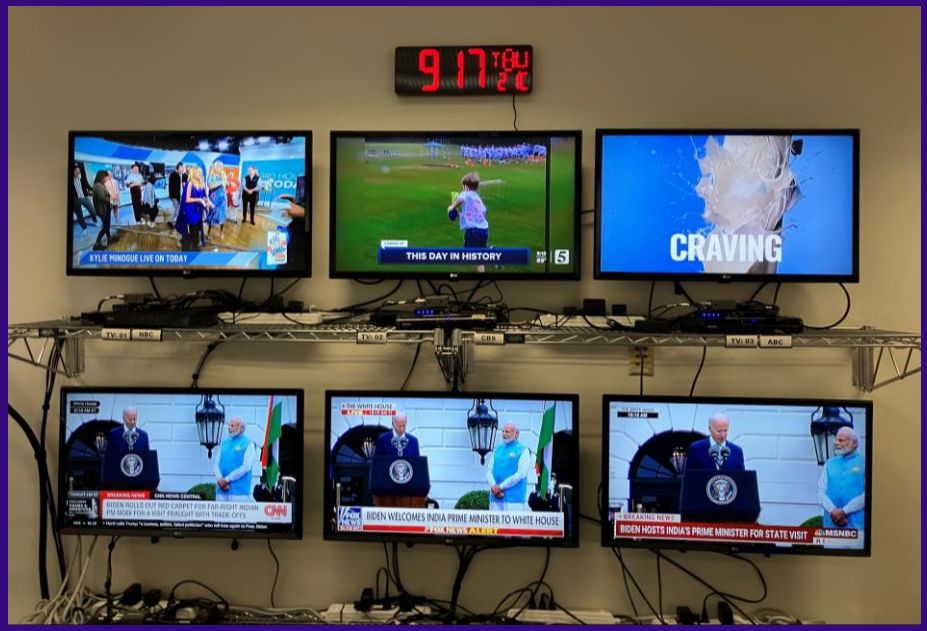
© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Challenge: Automation

The TV News Archive requires a highly reliable recording system that records the nightly news on five channels and maintains a 40-day buffer of television to monitor breaking news events.

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Automated Recording



© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Lambda

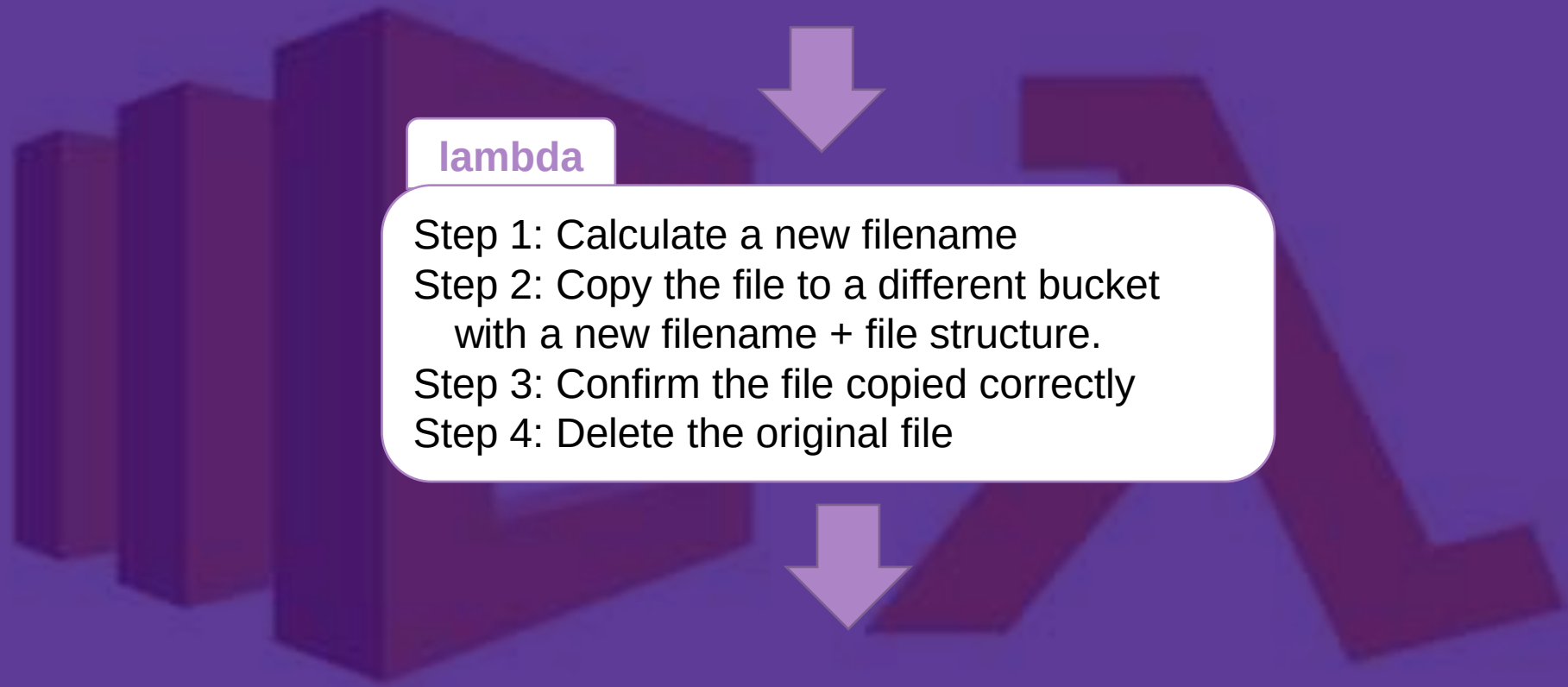
Runs code without servers or clusters

1 million requests per month for free

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Executes a snippet of code

```
New File = "VID_20221207T082900.mp4"
```



lambda

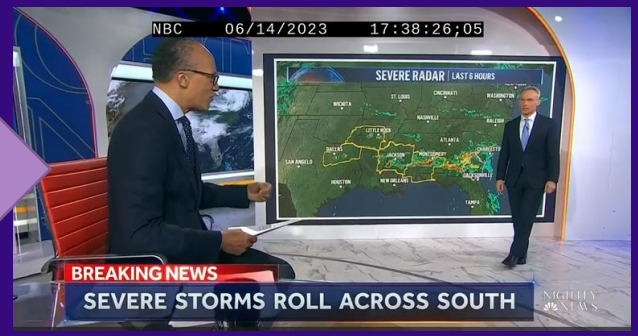
Step 1: Calculate a new filename
Step 2: Copy the file to a different bucket
with a new filename + file structure.
Step 3: Confirm the file copied correctly
Step 4: Delete the original file

```
Different Name & Location = "2022/12/VID20221207.mp4"
```

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

MediaConvert

Type: Primary Archive Version
Bitrate: 6 Mbps



Type: Proxy Access Version
Bitrate: 3 Mbps
Network, Date, Timestamp

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Lambda + MediaConvert

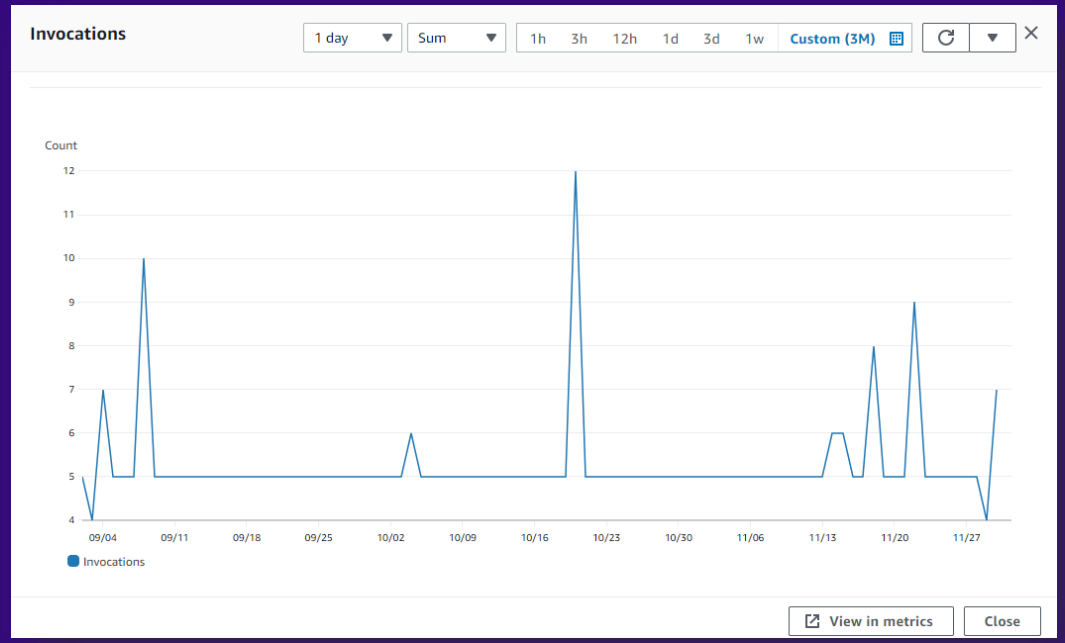
```
New File = "ABC_20190807T222901.mp4"
```

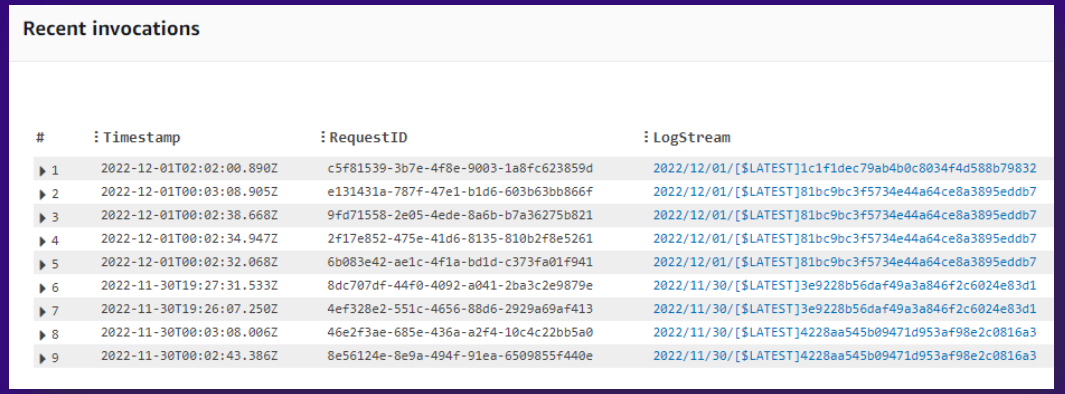
lambda

Step 1: Calculate Network, Date
Step 2: Convert Start time to Central Time
Step 3: Prepare settings for transcode job
Step 4: Start transcode job with watermark

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Dashboards & Logs





© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Challenge: Transcripts & Captions

While summaries are helpful, many forms of research require a full account of what was said on TV. Closed captions are also essential for universal access.

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

62,000 hours transcribed in 2 months!

SPLIT THE COLLECTION BY PRESIDENTIAL ADMINISTRATION, THEN FOR EACH GROUP:

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Challenges Addressed

- Automation
- Storage
- Preservation
- Enhanced research

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Jim Duran

Director
Vanderbilt Television News Archive
Vanderbilt University Libraries
jimduran@vanderbilt.edu
tvnews.Vanderbilt.edu

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.