

Samvera and AWS

Samvera Connect
November 2017

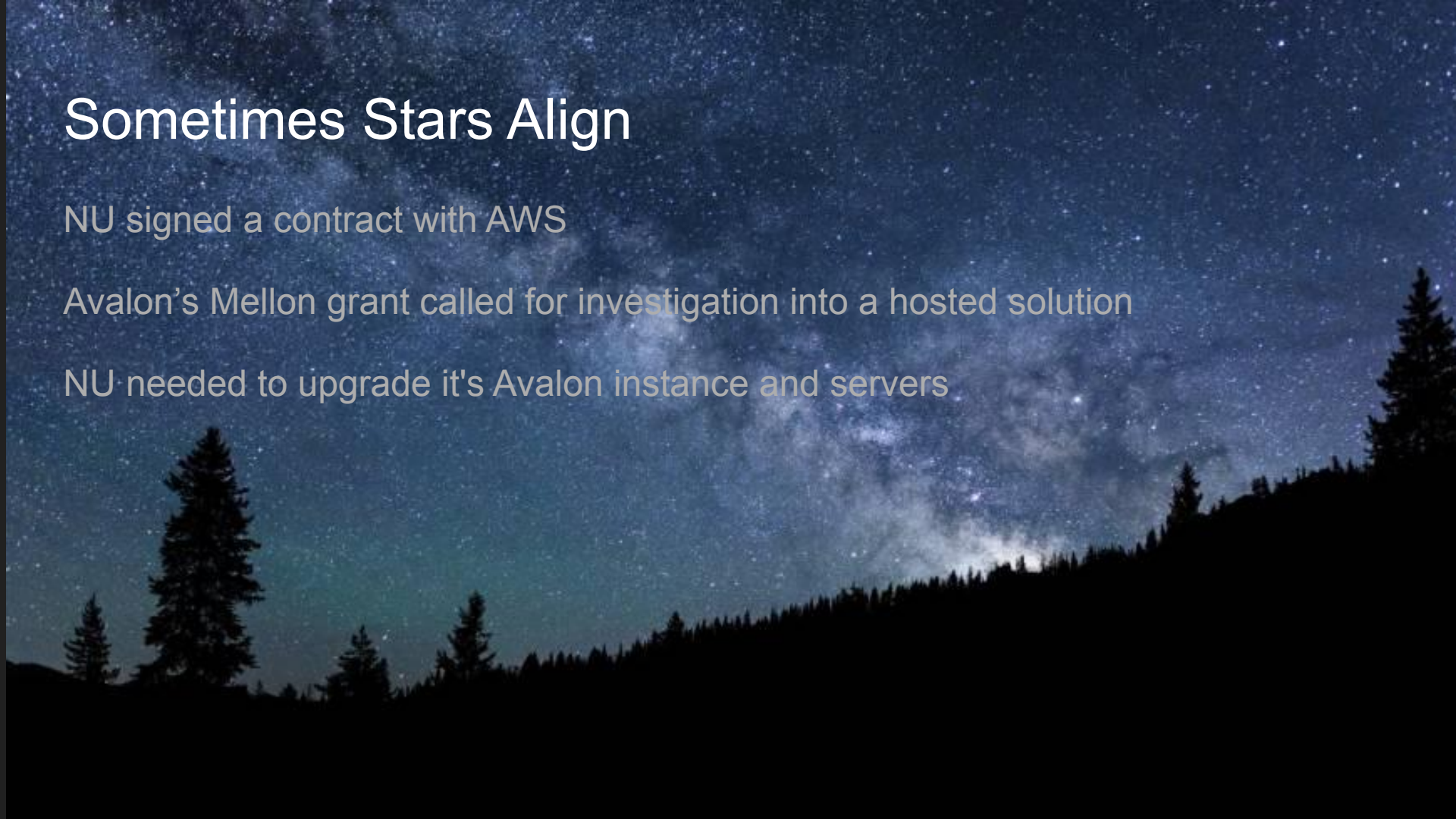
Northwestern
University

Sometimes Stars Align

NU signed a contract with AWS

Avalon's Mellon grant called for investigation into a hosted solution

NU needed to upgrade it's Avalon instance and servers





Avalon is complex, so it was a good candidate

Streaming

Transcoding

Large-scale storage

And Fedora, Solr, Web Front Ends, Workers, Etc.

How we pitched it

Simplified Development

Better user experience (global streaming, faster transcoding)

Infrastructure as code (staging, dev, prod are identical)

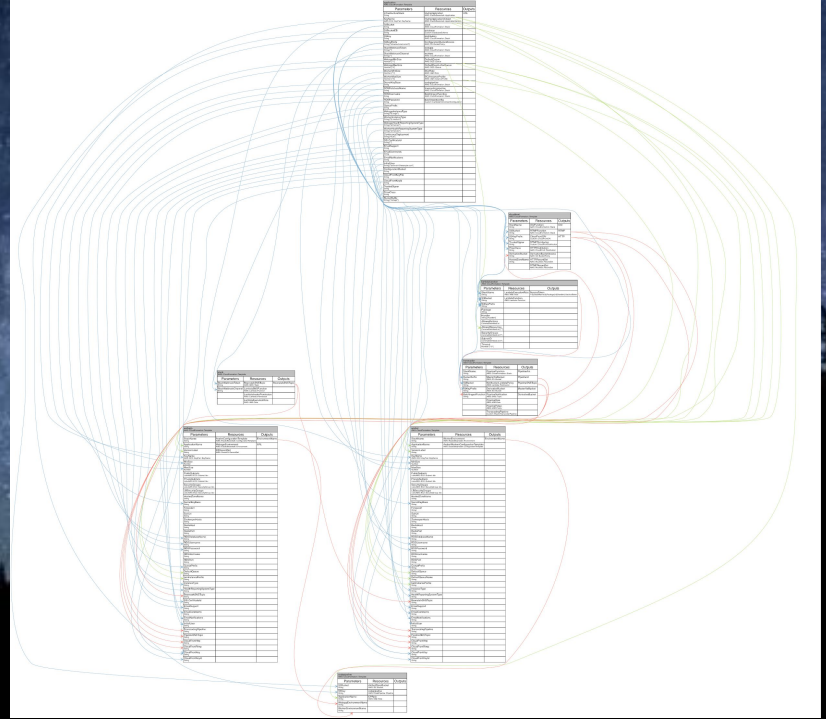
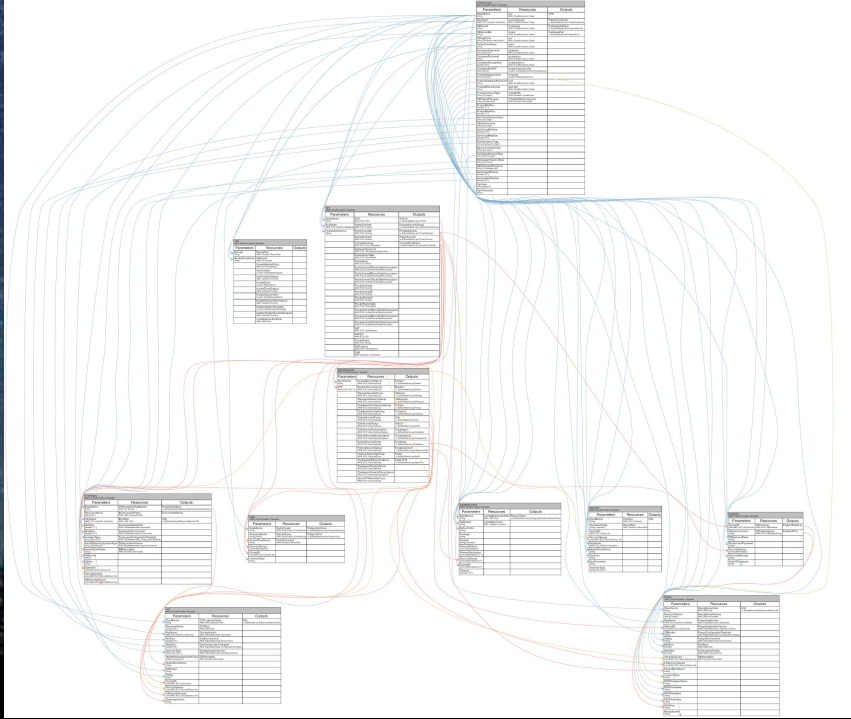
Reasonable and transparent costs

From Local Servers to Cloud Infrastructure

- Structural considerations
- Technical Challenges

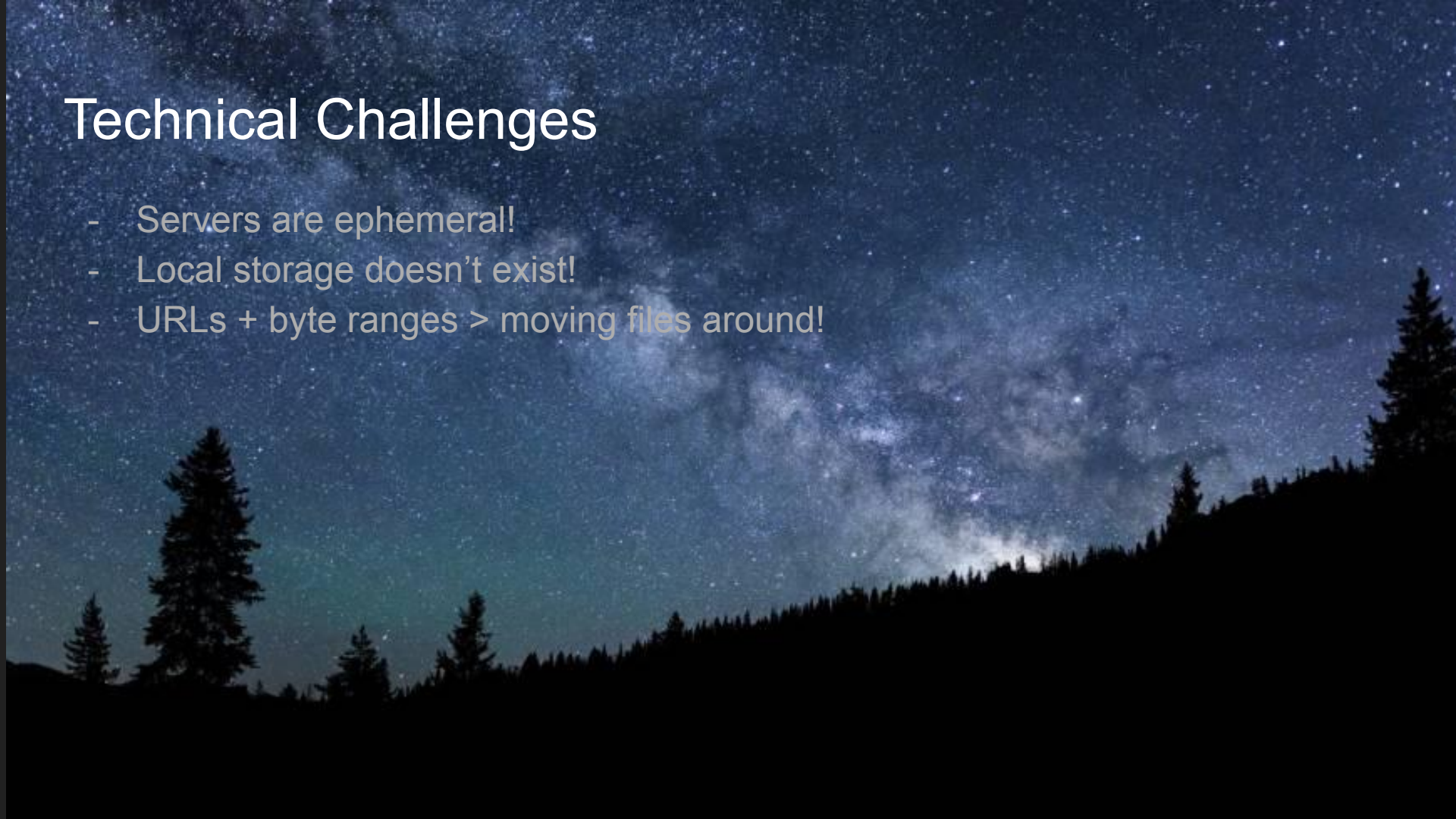


Structural Considerations



Technical Challenges

- Servers are ephemeral!
- Local storage doesn't exist!
- URLs + byte ranges > moving files around!



Example: Avalon File Characterization

The Old Way

- Write uploaded file (or download referenced file) to local storage
- Shell out to MediaInfo with path to local file
- Characterization tool has access to the full file on disk.

The New Way

- Have user POST file to one-off direct browser-to-S3 upload URL (or transfer referenced file to S3)
- Generate an HTTP URL pointing to the file in the S3 bucket
- Shell out to MediaInfo with the S3 URL
- Characterization tool uses byte-range requests to retrieve only the parts of the file it needs.

Monthly costs by service

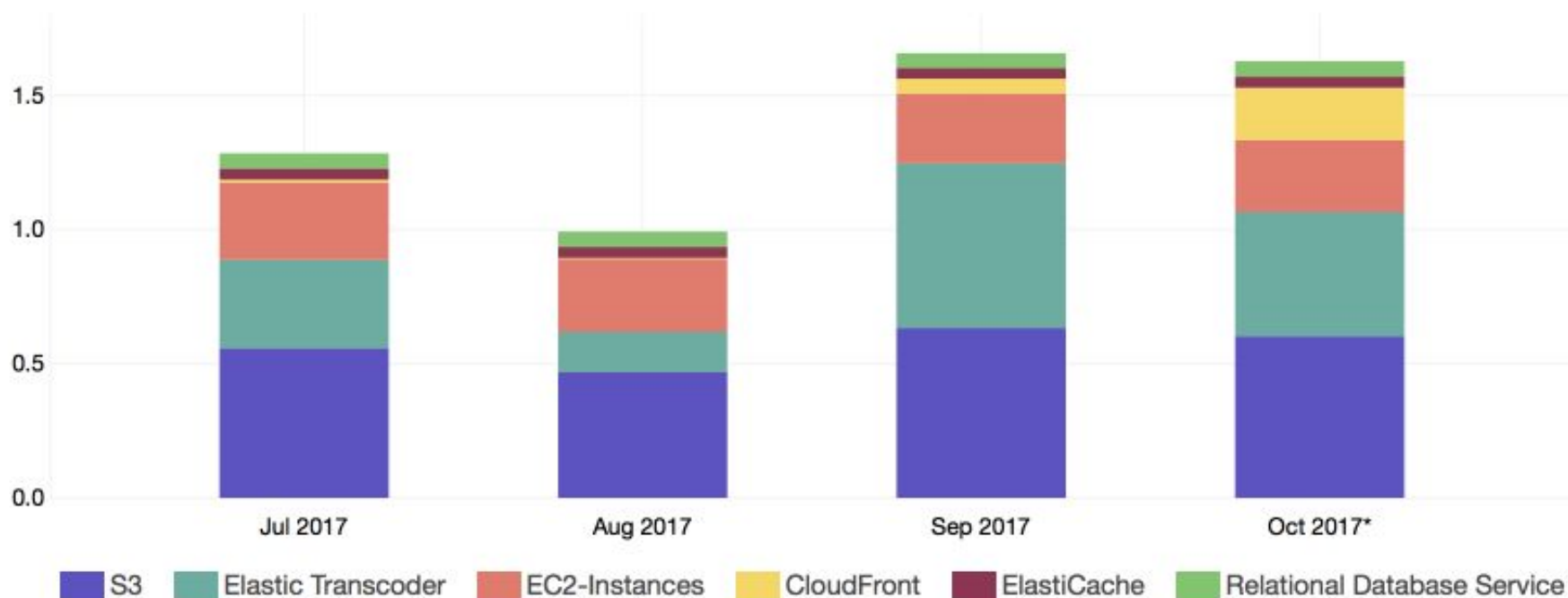
Jul 01, 2017 - Oct 31, 2017

Monthly

Group by: Service

Stack

Costs (\$ in thousands)





THE STANFORDS

PRODUCT OWNERSHIP



DevOps



MATT GROENING



INTEGRATIONS & NOTIFICATIONS



DEPLOYMENT VALIDATION



EMBEDDED CONFIGURATION

Notre Dame

Cloud-First Initiative

- University wide initiative to move services into the "cloud"
 - Either by using Software-as-a-Service solutions
 - Or by using AWS services and VMs
- Announced in 2014
- Library started Cloud-First working group in Fall 2016

Library Website

- Moved main library website and content to AWS in Summer 2017
- The new site is very AWS native
 - Uses AWS serverless design (lambdas)
 - Serves content directly from S3
- Developers and System Administrators need to work closer together
- So many AWS services, difficult to choose the right one

Repository Migration

Preservation tape storage is still (and will be) on campus.

New Hyrax Application will be in AWS in containers

Undecided how auxiliary services will be structured

- Preservation storage will be local. But cached in S3?
- Batch ingest uses local NFS. Change to something else?