A Multi-Tenancy Cloud-Native Digital Library Platform

Yinlin Chen, Jim Tuttle, William A. Ingram

{ylchen, jim.tuttle, waingram}@vt.edu

Information Technologies and Services
Virginia Tech Libraries

VIRGINIA TECH

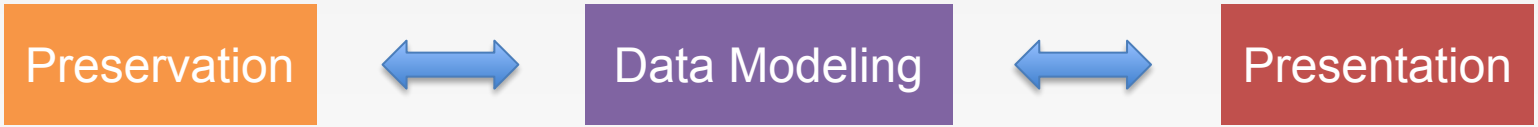
Agenda

- Cloud-native concept
- Virginia Tech Digital Library Platform (VTDLP)
- Design strategy
- Architecture overview
- Implementation overview
- VTL experiences

Cloud-native Concept

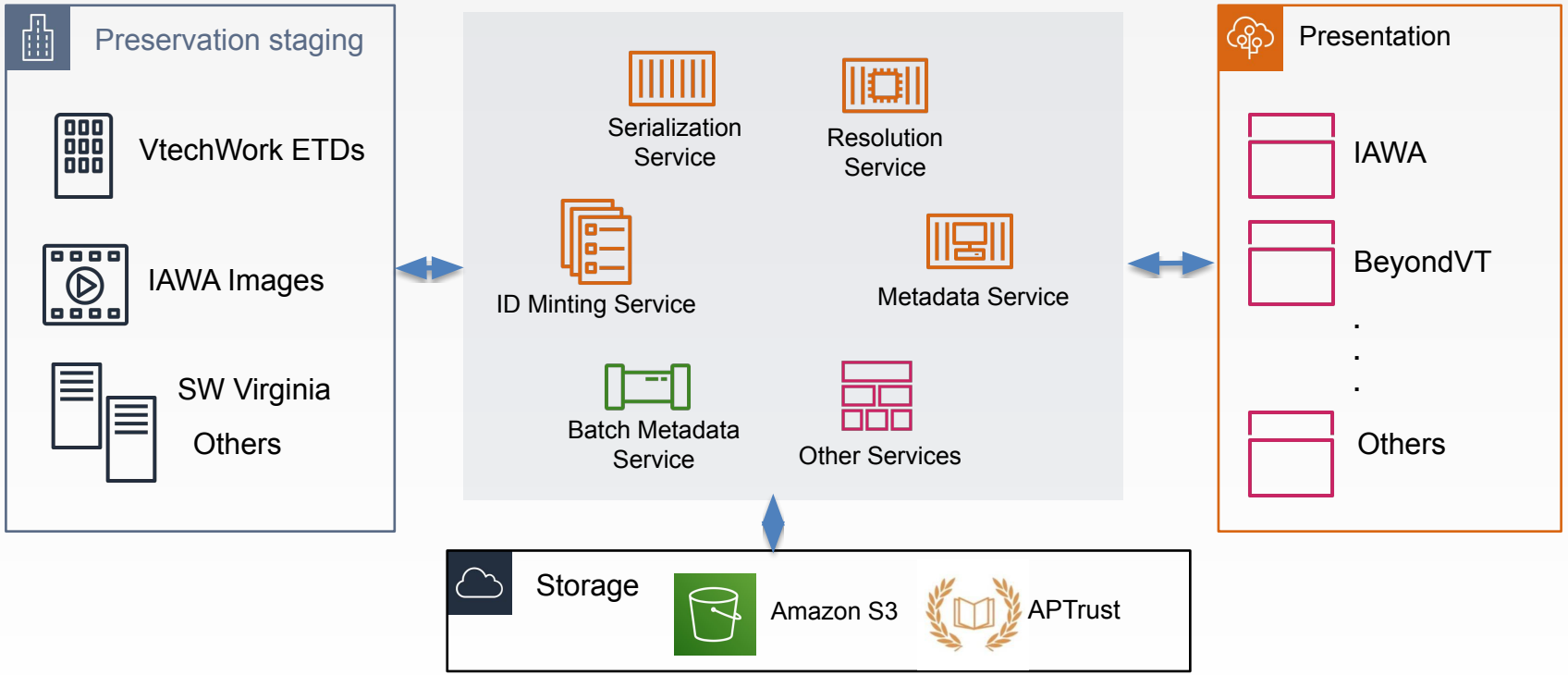
- Entire infrastructure is deployed in the Cloud (AWS)
- Platform is composed of a suite of microservices and managed services
- Focus on the business logic and workflow
- Utilize the advantages provided by the Cloud

Virginia Tech Digital Library Platform (VTDLP)



- New services to Digital Library Platform
  - ID Minting service, Access Service, Metadata service, …
- Migrating legacy services to Digital Library Platform
  - IAWA, VTechWork, …

VTDLP Overview



Design Strategy

- Cloud native (AWS ecosystem)
- Microservice/SOA (AWS lambda)
- Serverless (AWS managed services)
- CI/CD Pipeline
- Caching as much as possible
  - Static files
  - Lambda functions
- Automation as much as possible
  - Infrastructure as code
  - No manual provisioning or managing servers

AWS Ecosystem

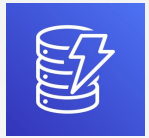

Amazon CloudFront


Amazon Route 53

Network & Content Delivery


AWS Lambda


Amazon EC2


Amazon DynamoDB


Amazon ES

Compute & Database


AWS Certificate Manager


Amazon SQS


Amazon SNS

Messaging


IAM


AWS Organizations

Security & Identity


Amazon S3


Amazon Glacier

Storage


AWS CLI


AWS CloudFormation


AWS CloudTrail


Amazon CloudWatch


AWS Amplify


Amazon Pinpoint

Management

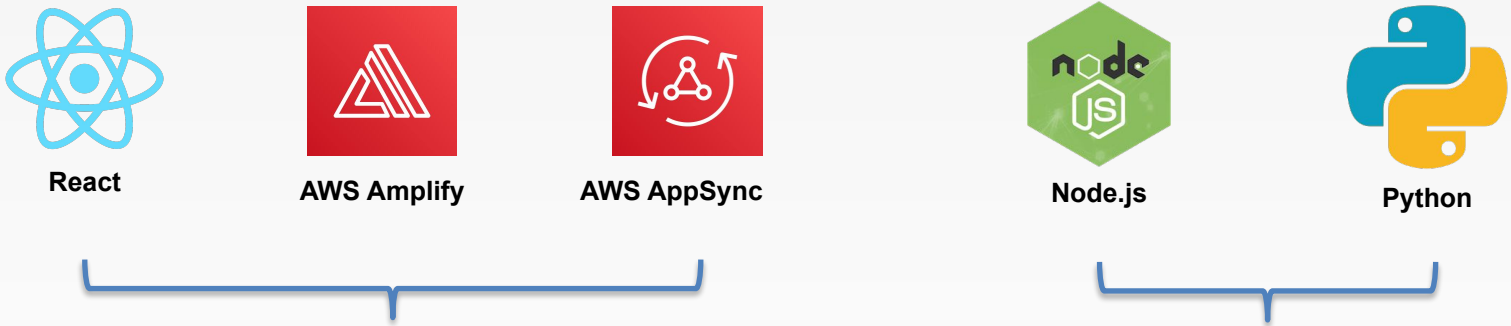

Amazon API Gateway


Amazon Cognito

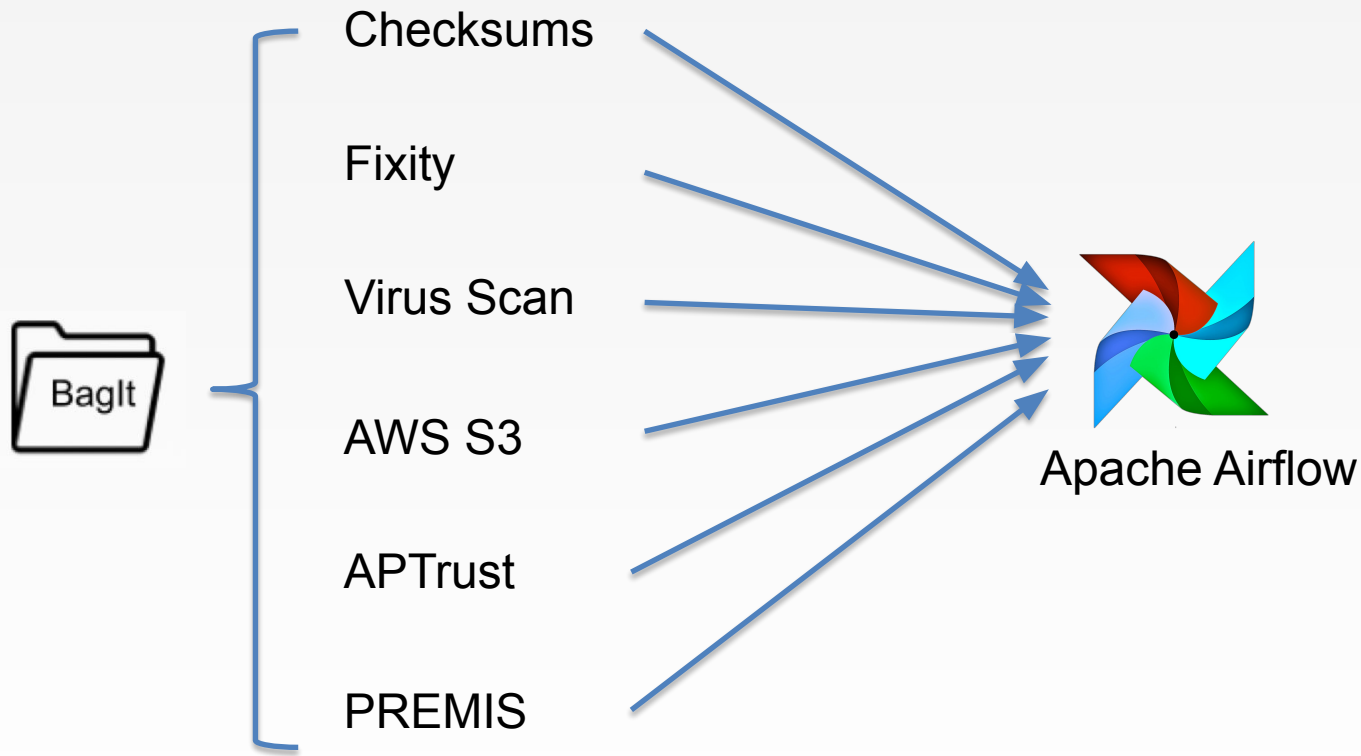
Services

VIRGINIA TECH

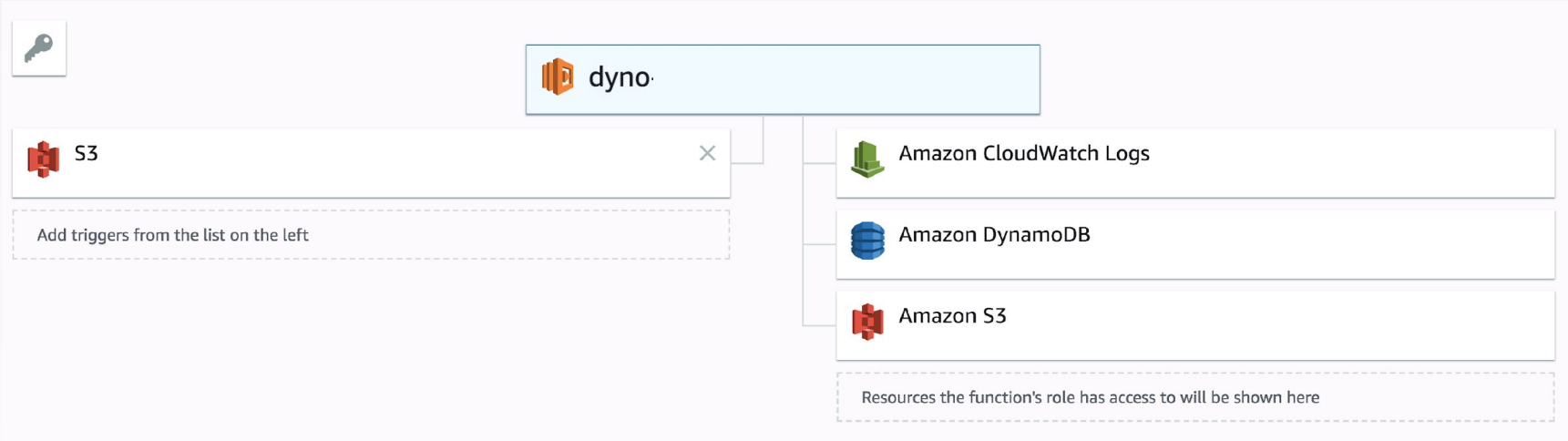
Software stacks



Preservation Pipeline

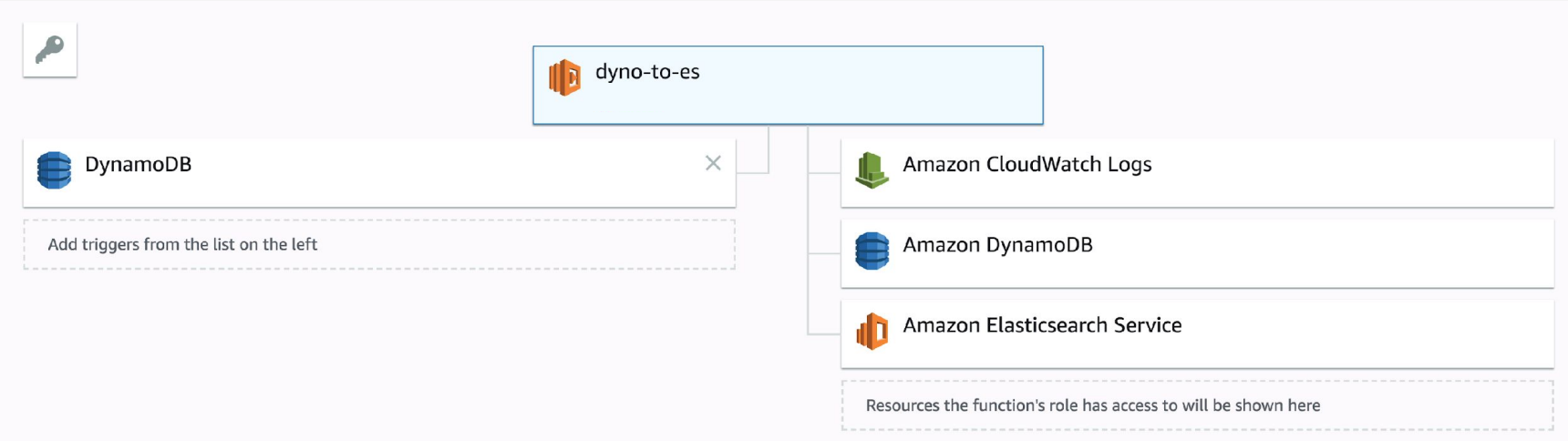


Lambda Example – Metadata file



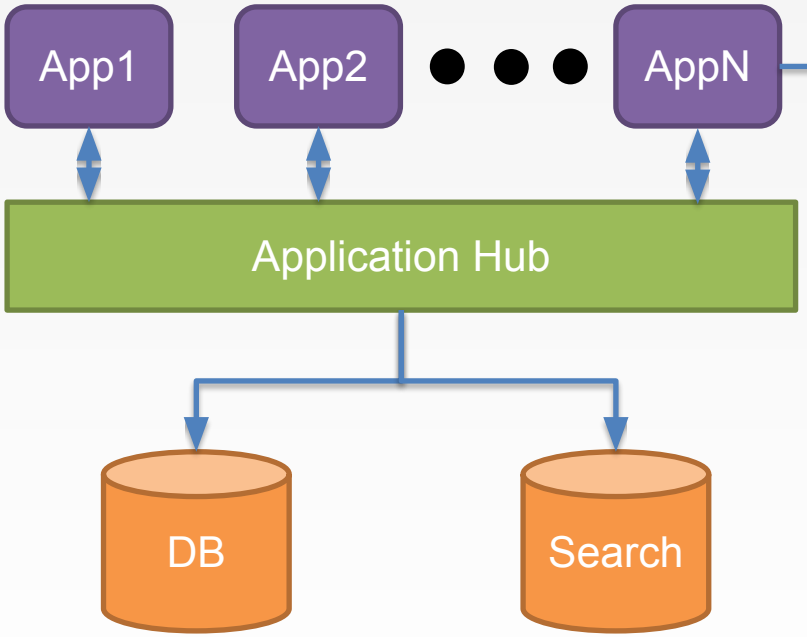
1. File upload to S3
2. S3 triggers a Lambda function
3. Lambda function parses file content and inserts/updates record in the DynamoDB

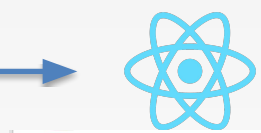
Lambda Example – DynamoDB / ES

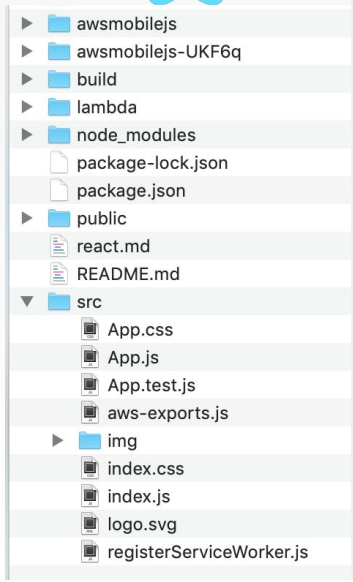


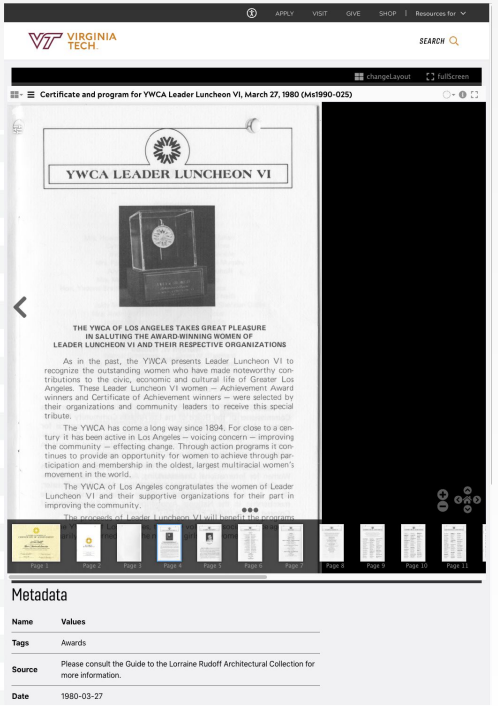
1. Data modifications in DynamoDB will trigger a Lambda function
2. Lambda function captures changes and updates Amazon ES

Presentation - Multi-Tenant Architecture

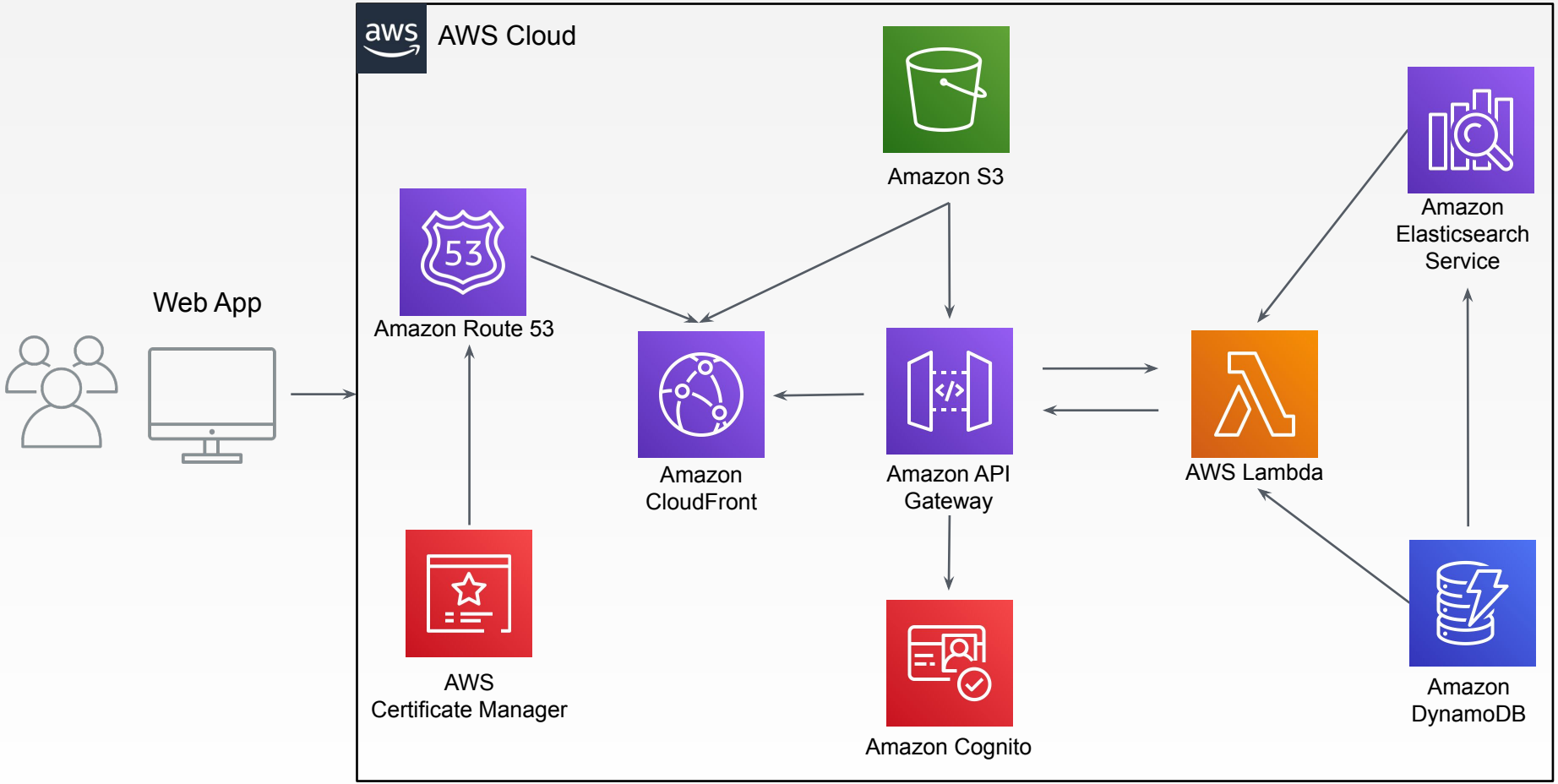








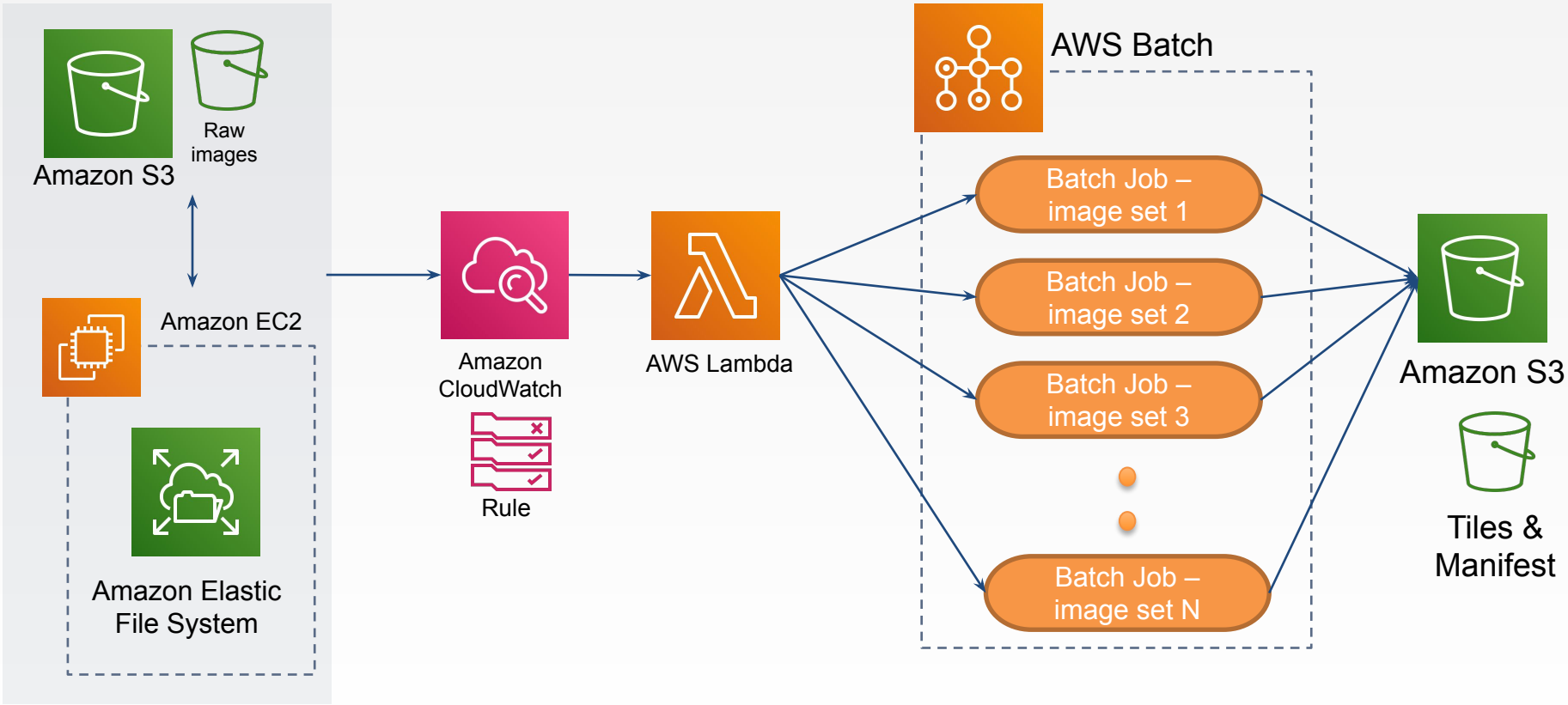
VIRGINIA TECH


AWS Cloud
Amazon S3
Amazon Elasticsearch Service
Web App
Amazon Route 53
Amazon CloudFront
Amazon API Gateway
AWS Lambda
AWS Certificate Manager
Amazon Cognito
Amazon DynamoDB

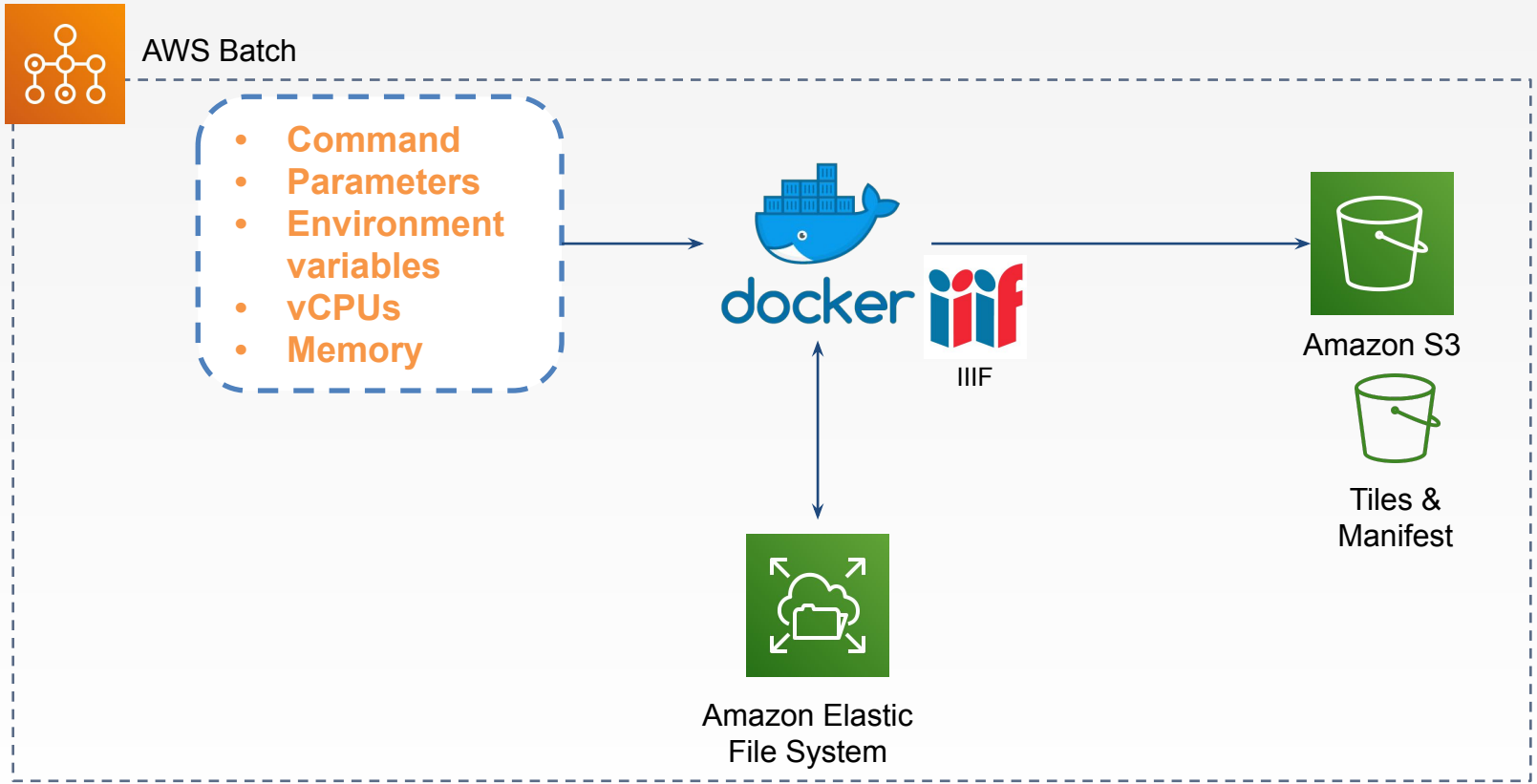
The International Archive of Women in Architecture

- A level 0 compliant image server using Amazon S3 and Amazon CloudFront
- Tiles images, manifest JSON files, and etc.
- Terabytes of scan images to be processed

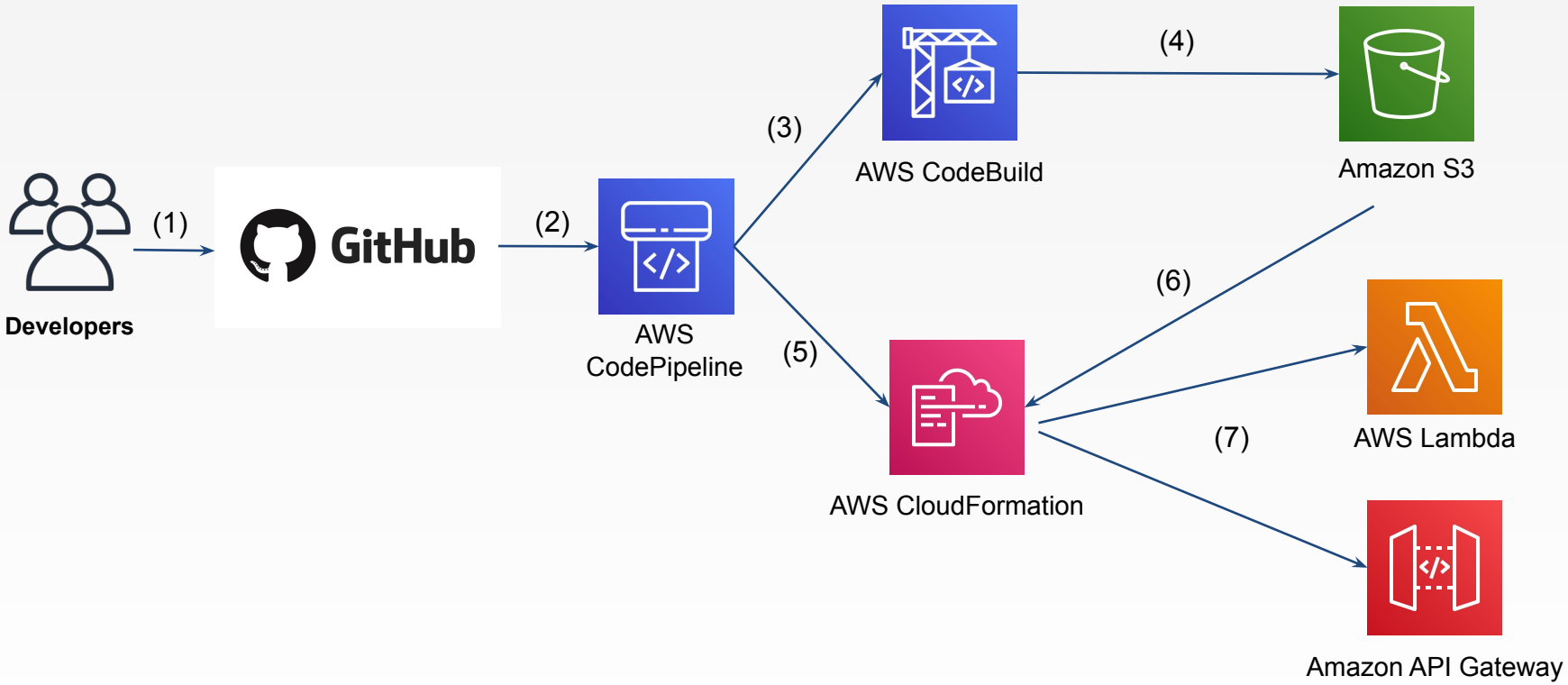
Image processing workflow



Batch job - IIIF_S3 Docker



CI/CD with AWS



Cloud benefit - Backup examples

- S3
  - Amazon S3 is 99.999999999% durability and 99.99% availability.
  - On average, may lose one of 10,000 objects every 10 million years or so.
  - Cross-region replication
- DynamoDB
  - Point-in-time recovery (Last 35 days)
  - On-Demand Backup (Stored in S3)
- ElasticSearch
  - Daily snapshots (Last 14 days)
  - On-Demand Backup (Stored in S3)

VTL Experiences

- Entire development team is AWS certified

- One AWS Certification Subject Matter Expert (SME)
- AWS trainings and conferences
- Thinking and implementing new ideas the Cloud way

Q & A

Thank You!