Virginia Tech Libraries' Next Gen Digital Libraries Platform

Yinlin Chen and James Tuttle

{ylchen, james.tuttle}@vt.edu

Virginia Tech Libraries

VIRGINIA TECH

Agenda

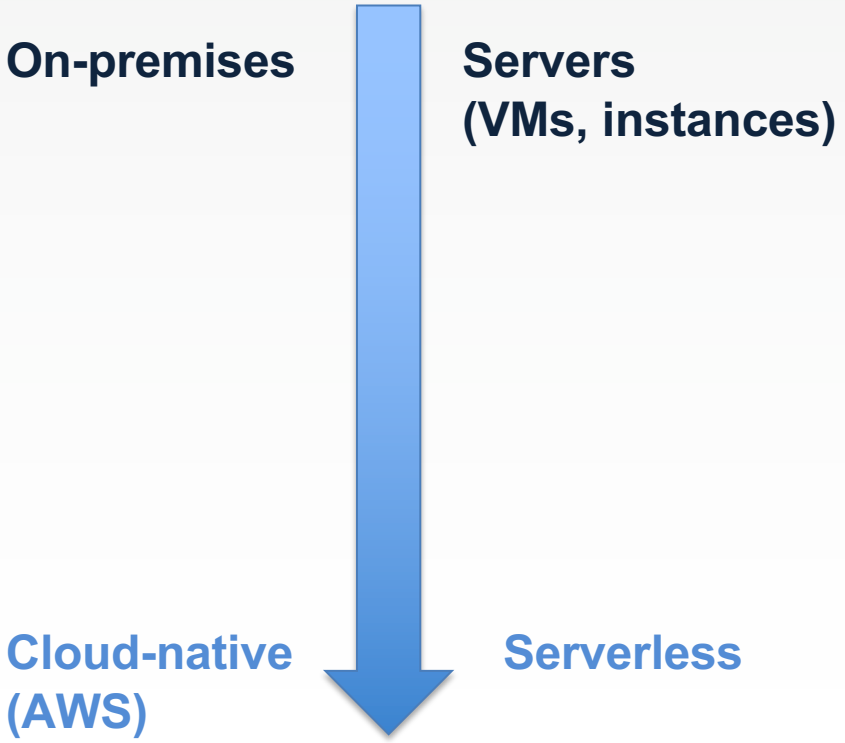
- Problem Space
- DLD Projects
- Cloud-native
- Serverless & Microservices
- Virginia Tech Digital Library Platform (VTDLP)
- Architecture Overview
- Outcome
- Next Steps

Problem Space

- Numerous, web applications with similar stacks stretching resources
- Limited in-house capacity to address performance, resilience, and scaling
- Library-specific software requires training or competing for few experienced library devs

DLD Projects

- FishTraits database
- ETDplus
- VTechData
- GeoData
- CollabVT
- Fedora
- VTDLP
- IAWA
- ……

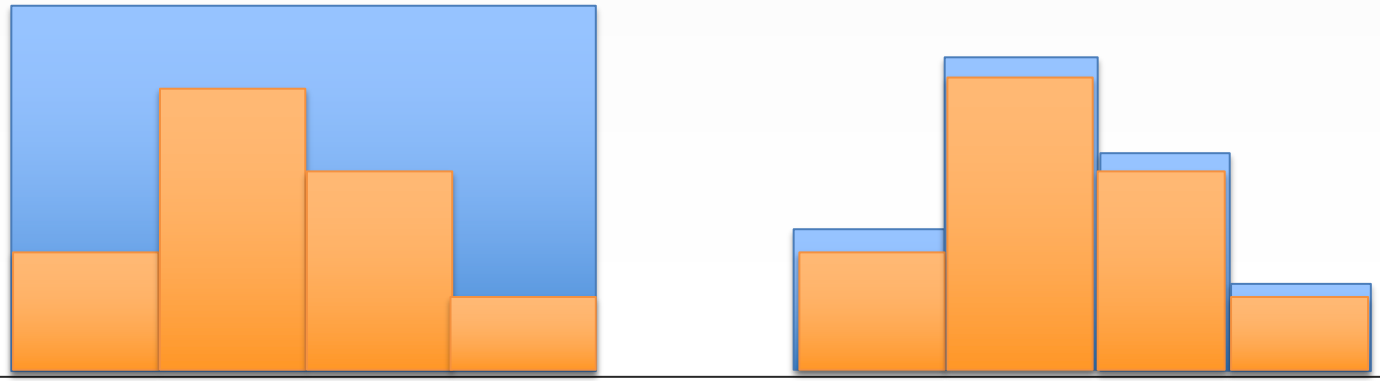


Cloud Native

- 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
  - fault-tolerant, auto-scale, update/rollback without downtime, etc.
- Facilitate the development process
- Optimize resource utilization
- Optimize and reduce cost

Resource Usage Optimization and Automation

- Consume only the required resources for the applications
- Scale up and down automatically
- Service and function oriented, not server oriented
- Utilize cloud services to help understand applications (CloudWatch, Auto Scaling, Trusted Advisor, etc.)



VIRGINIA TECH

Serverless

Does not mean “There are no servers at all”.

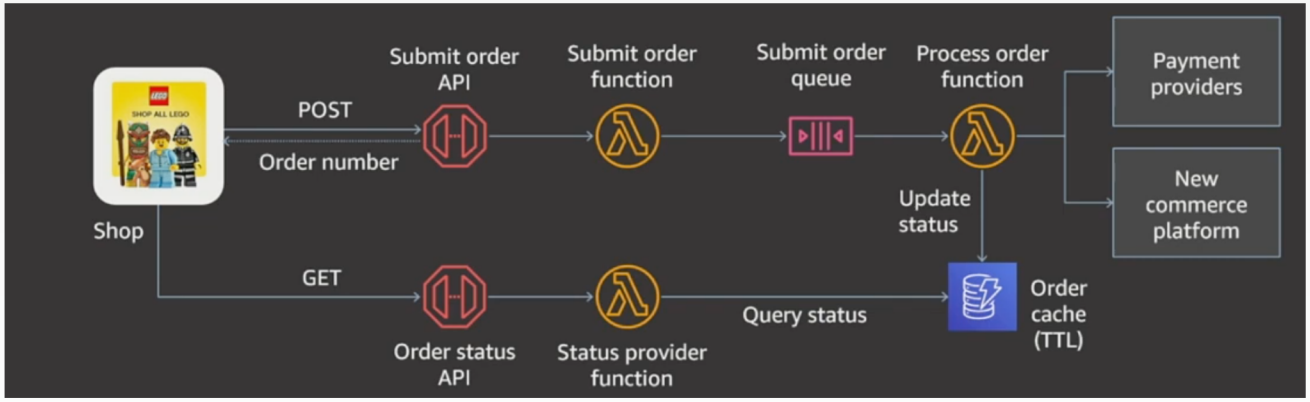
Does mean “Use fully managed services”.

Focus on application development,
not server maintenance

Microservice

- Small applications that do one thing well
- Messaging enabled – communicate with messages
- Decentralized
  - Autonomously developed
  - Independently deployable
  - Can change independently of each service
  - Scale individually by load
- Built and released with automated processes
- More complex architecture

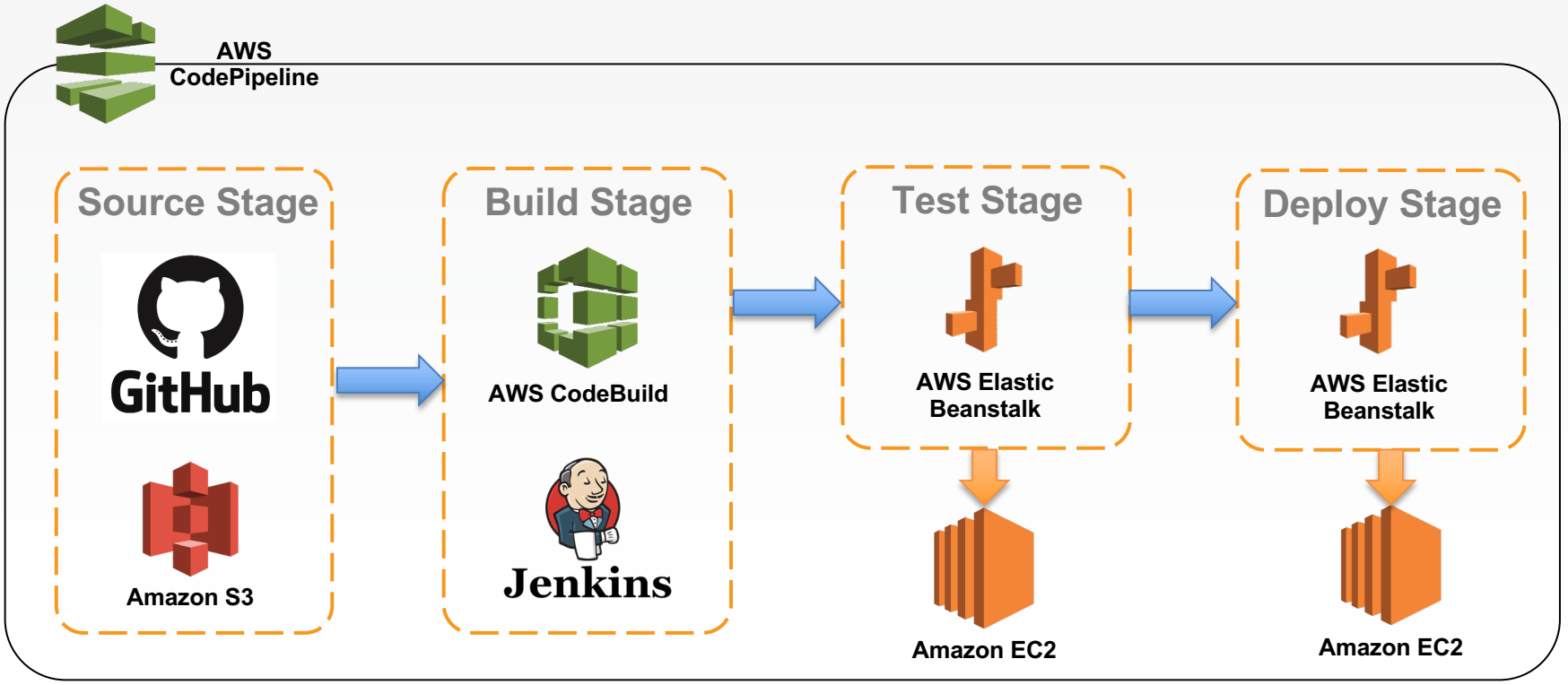
Shop.LEGO.com serverless on AWS



Images from [Lego AWS:reinvent 19 presentation]

VIRGINIA TECH

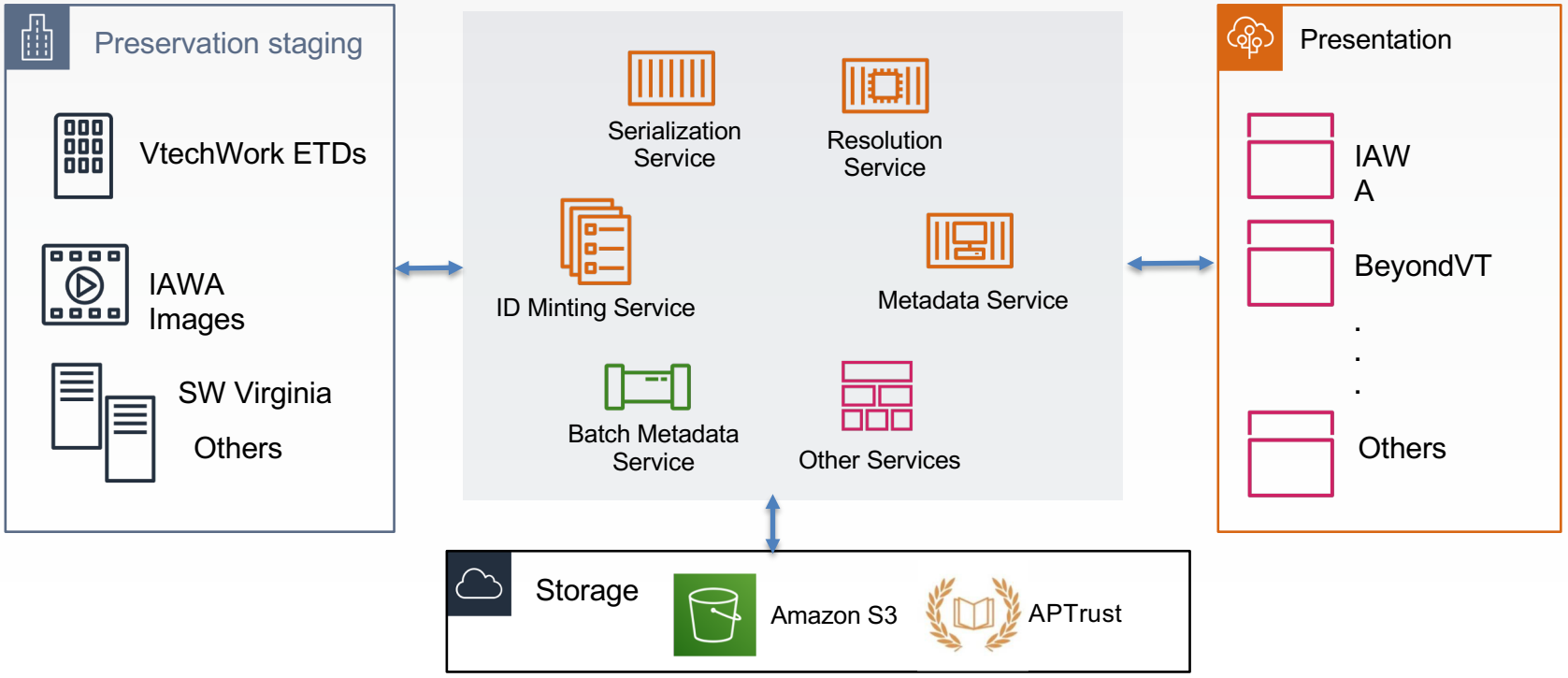
Continuous Integration and Delivery (CI / CD)



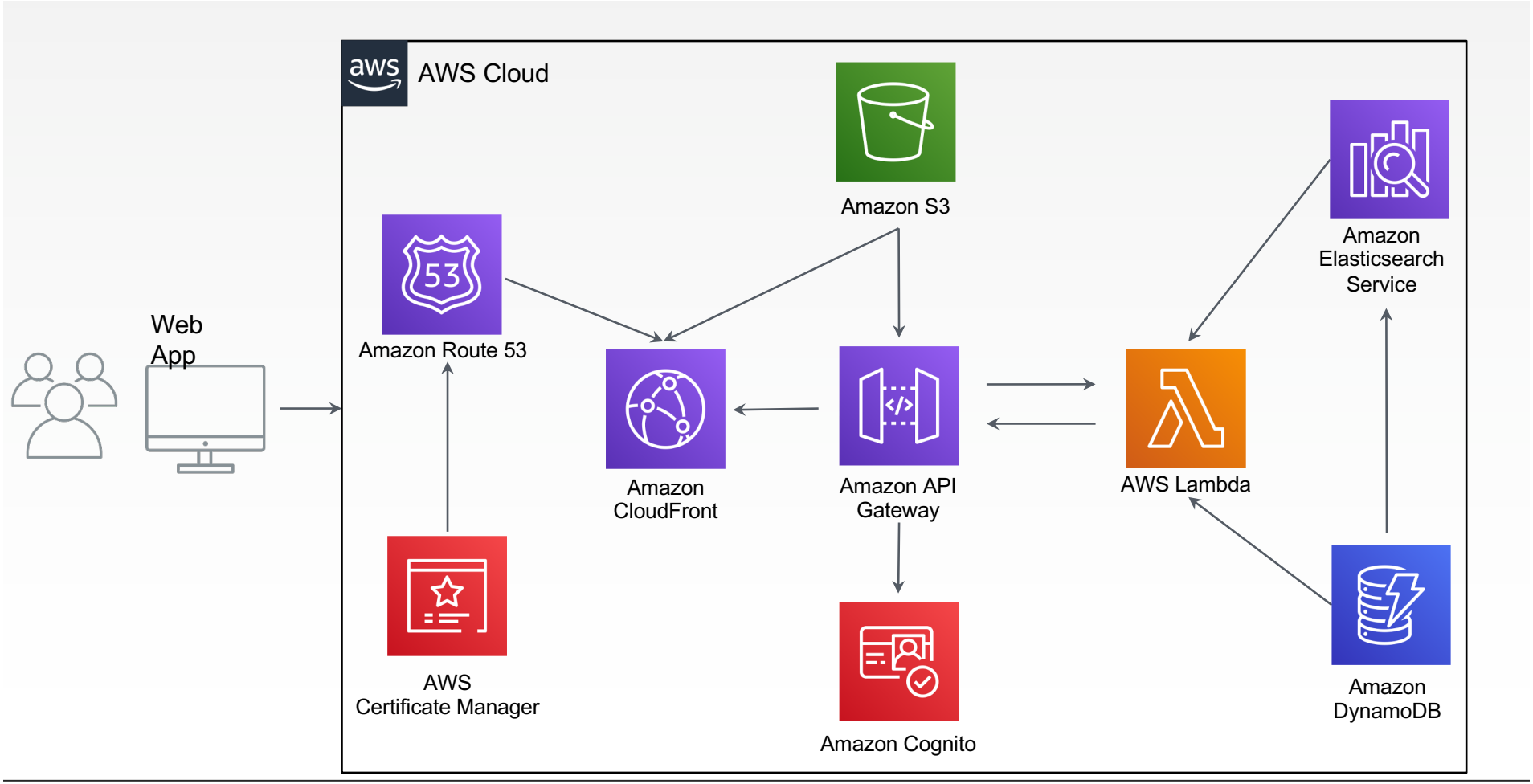
Virginia Tech Digital Library Platform (VTDLP)

- New services to Digital Library Platform
  - ID Minting service, Access Service, Metadata service, …
- Migrating legacy services to Digital Library Platform
  - IAWA, VTechWork, …
- A Multi-Tenancy Cloud-Native Digital Library Platform – OR 2019

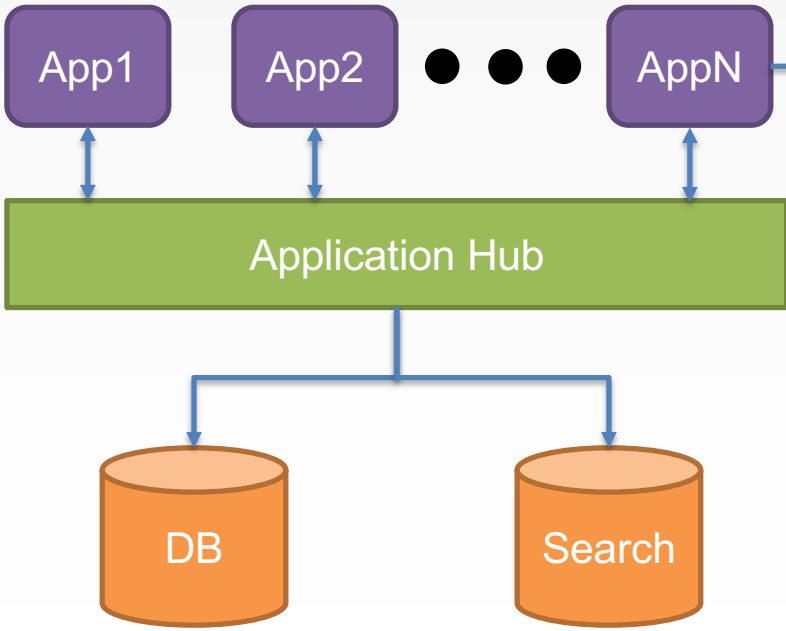
VTDLP Overview

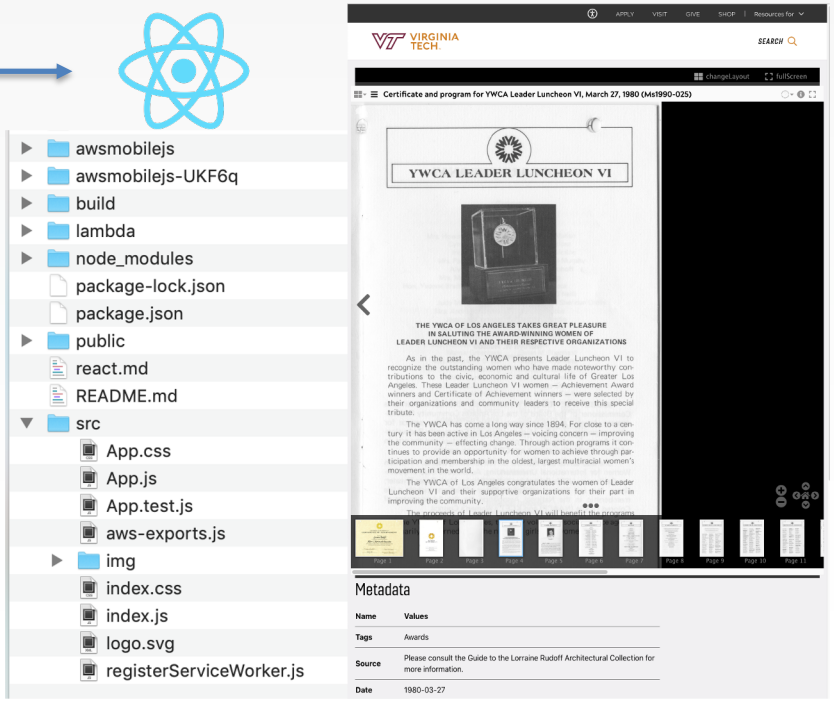


VIRGINIA TECH


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

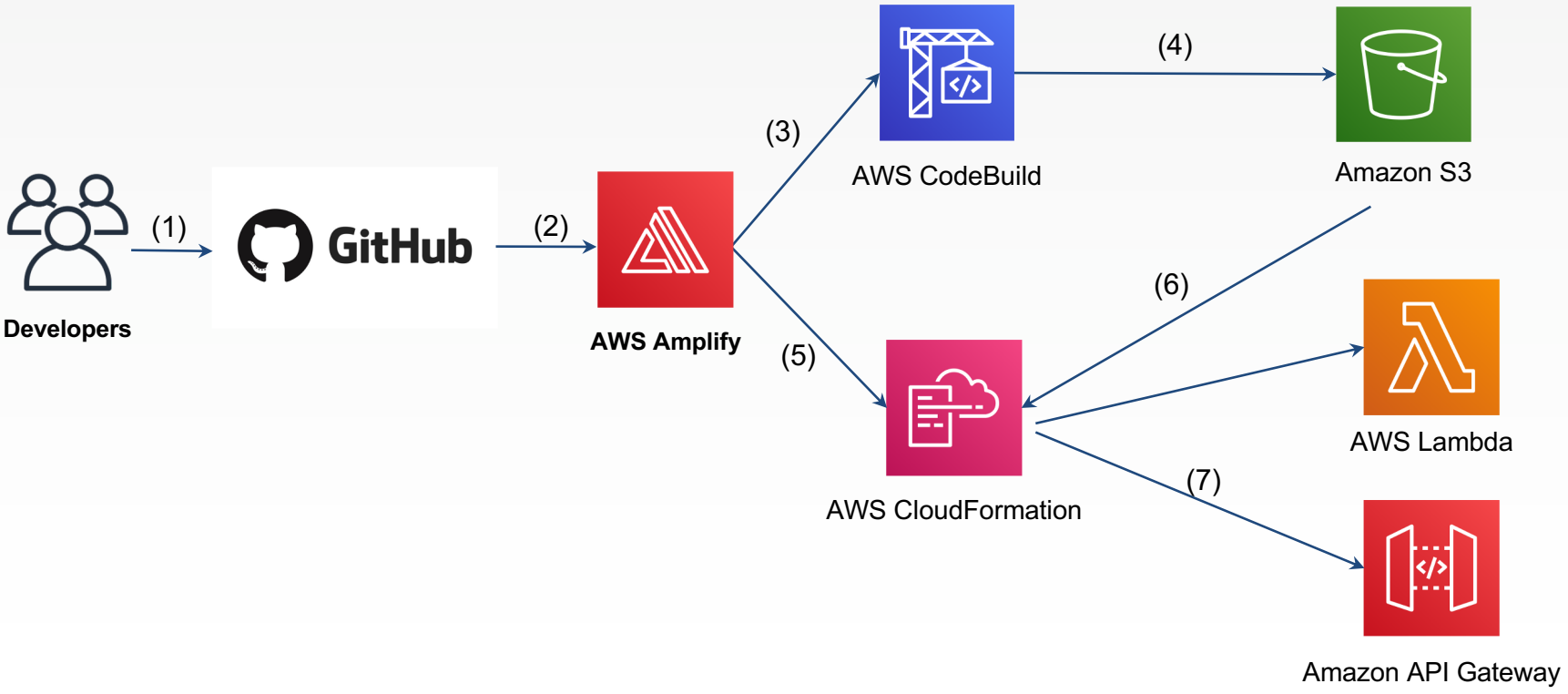
Presentation - Multi-Tenant Architecture



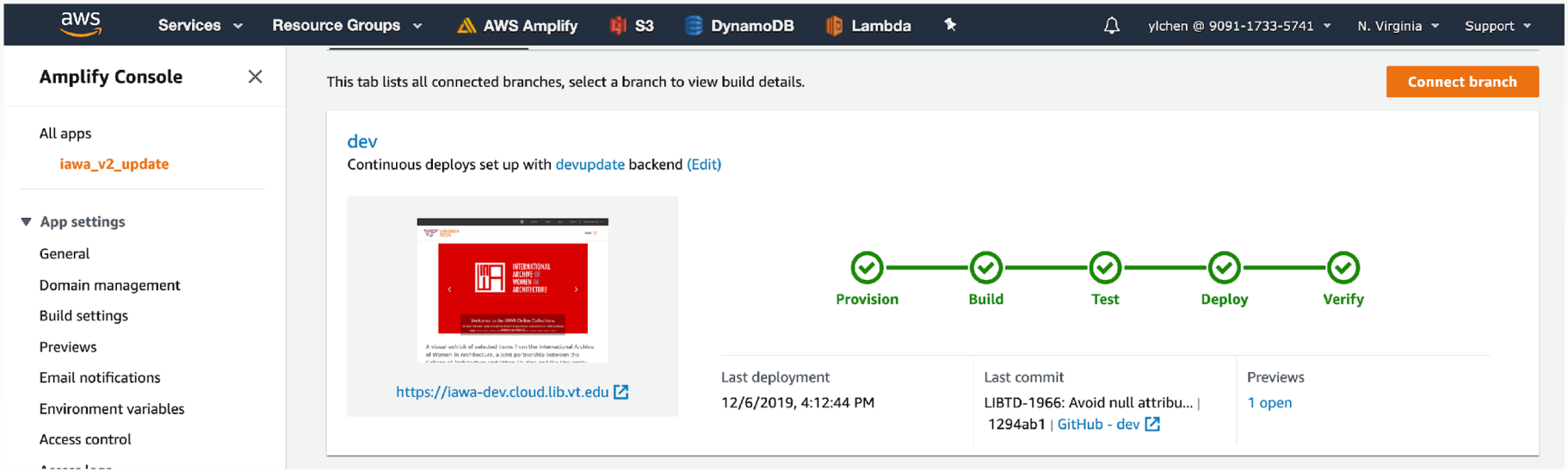


VIRGINIA TECH

CI/CD with AWS

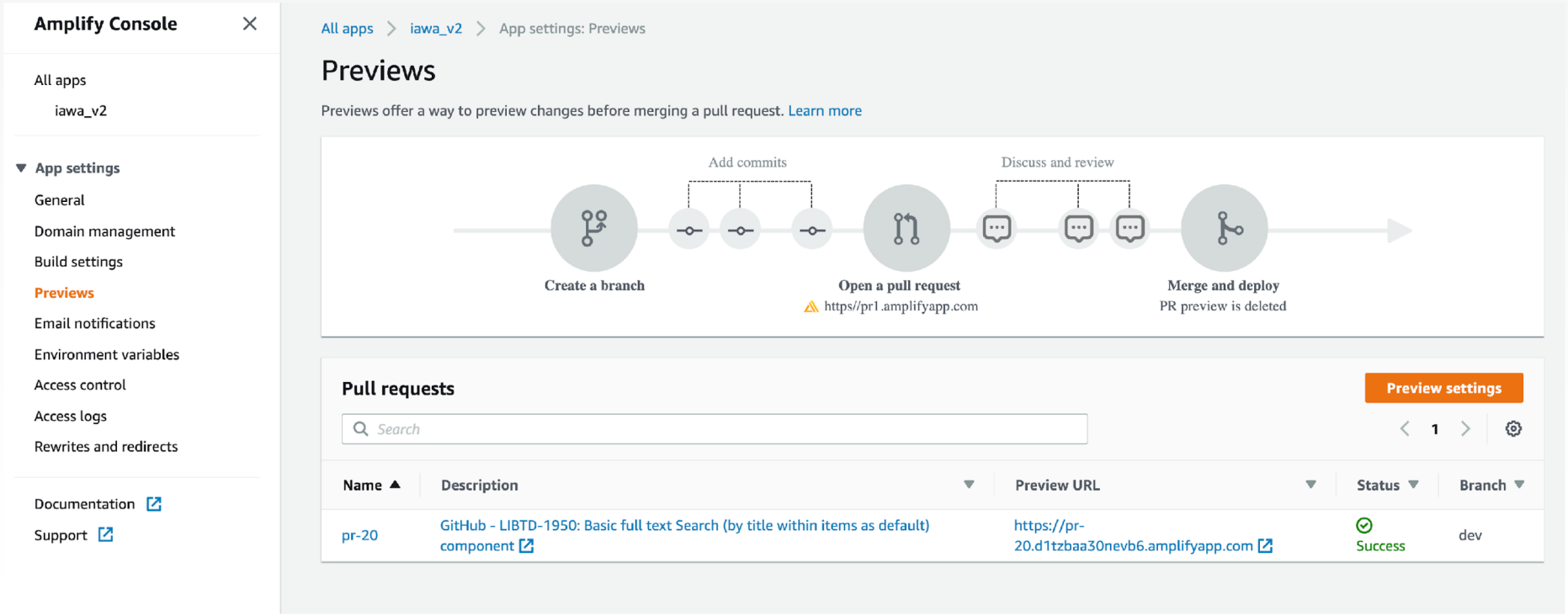


Automatic CI/CD Pipeline



VIRGINIA TECH

A New Version for each Pull Request

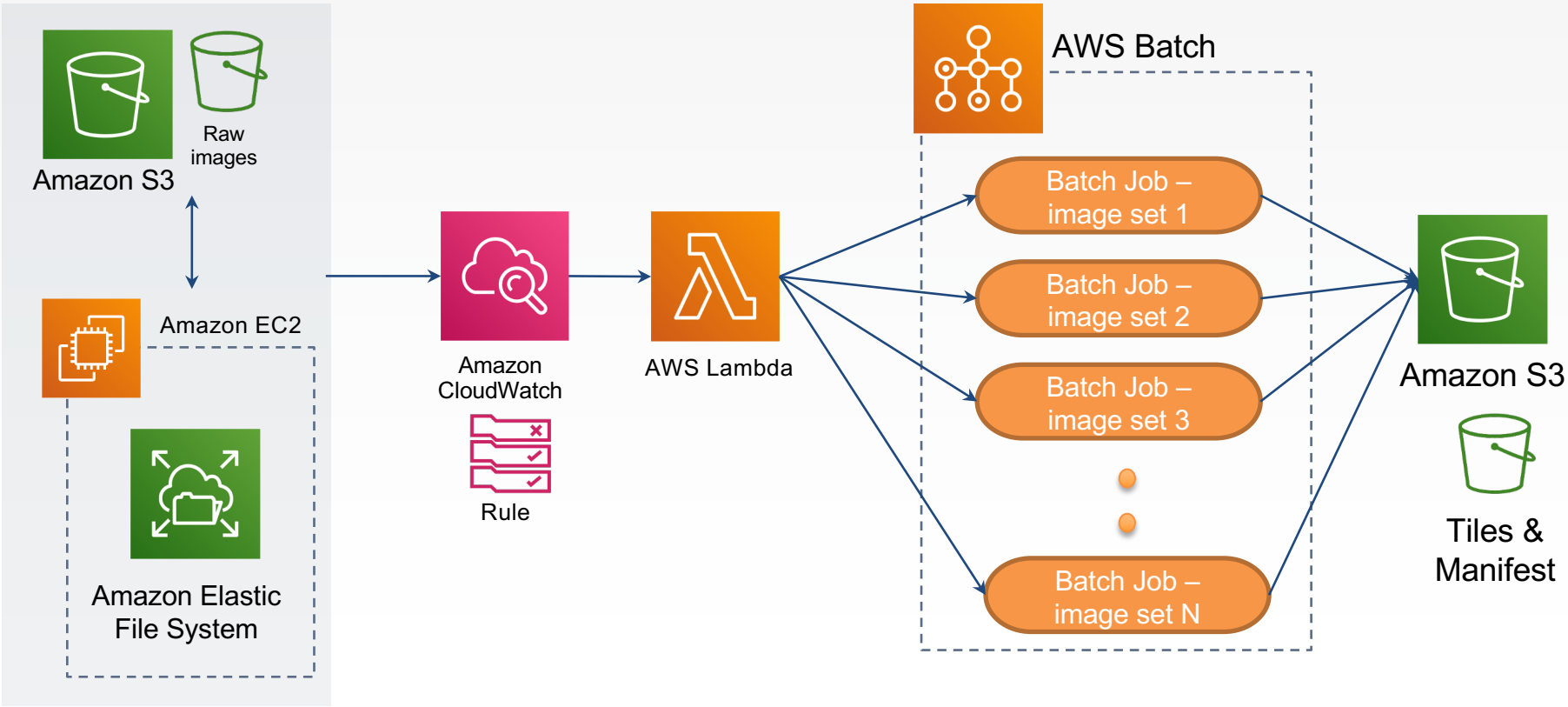


VIRGINIA TECH

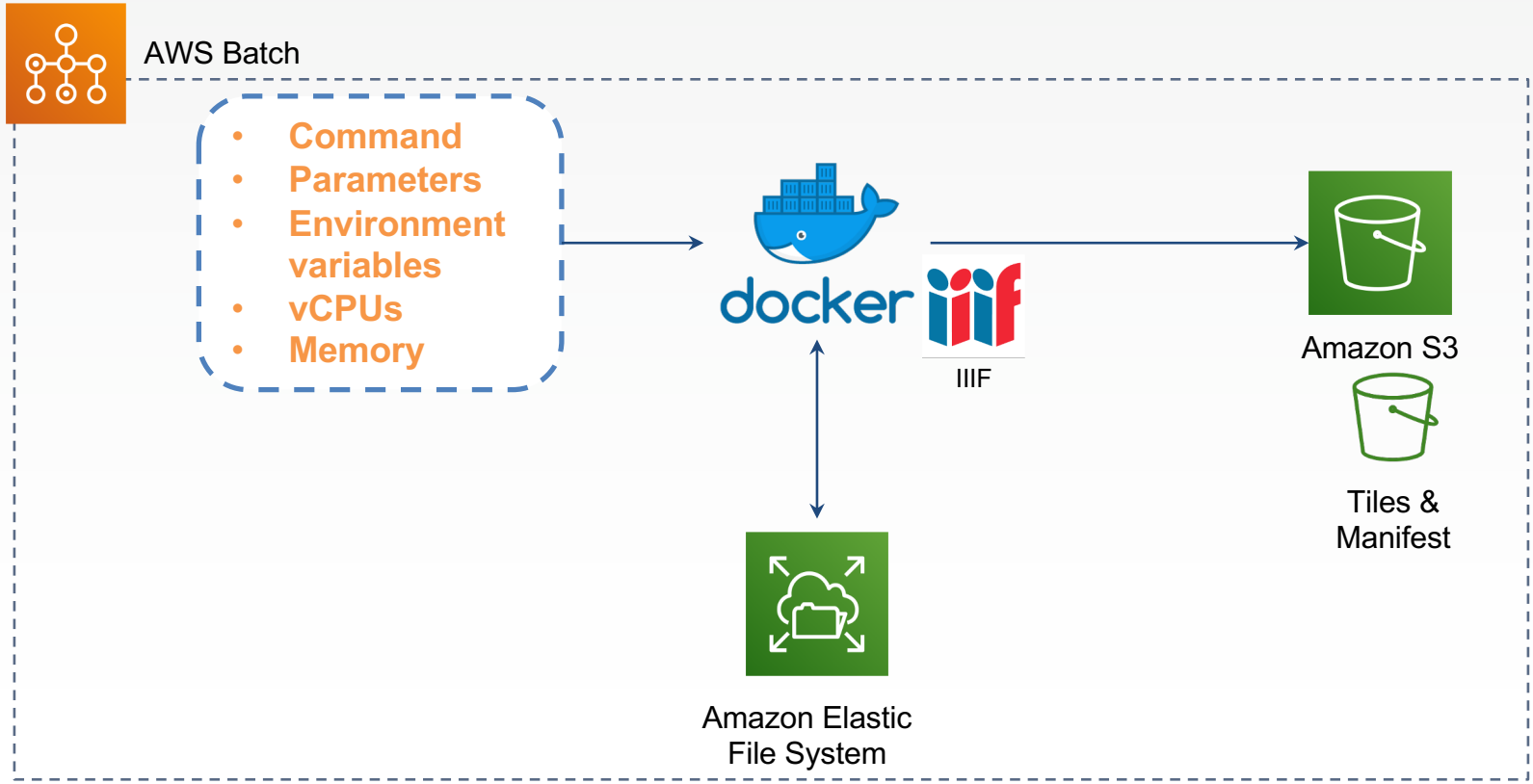
The International Archive of Women in Architecture

- 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
- Scaling IIIF image tiling in the cloud – Code4Lib Journal (To be published)

Image processing workflow



Batch job - IIIF_S3 Docker



Automatic Data Process Pipeline

Job queues

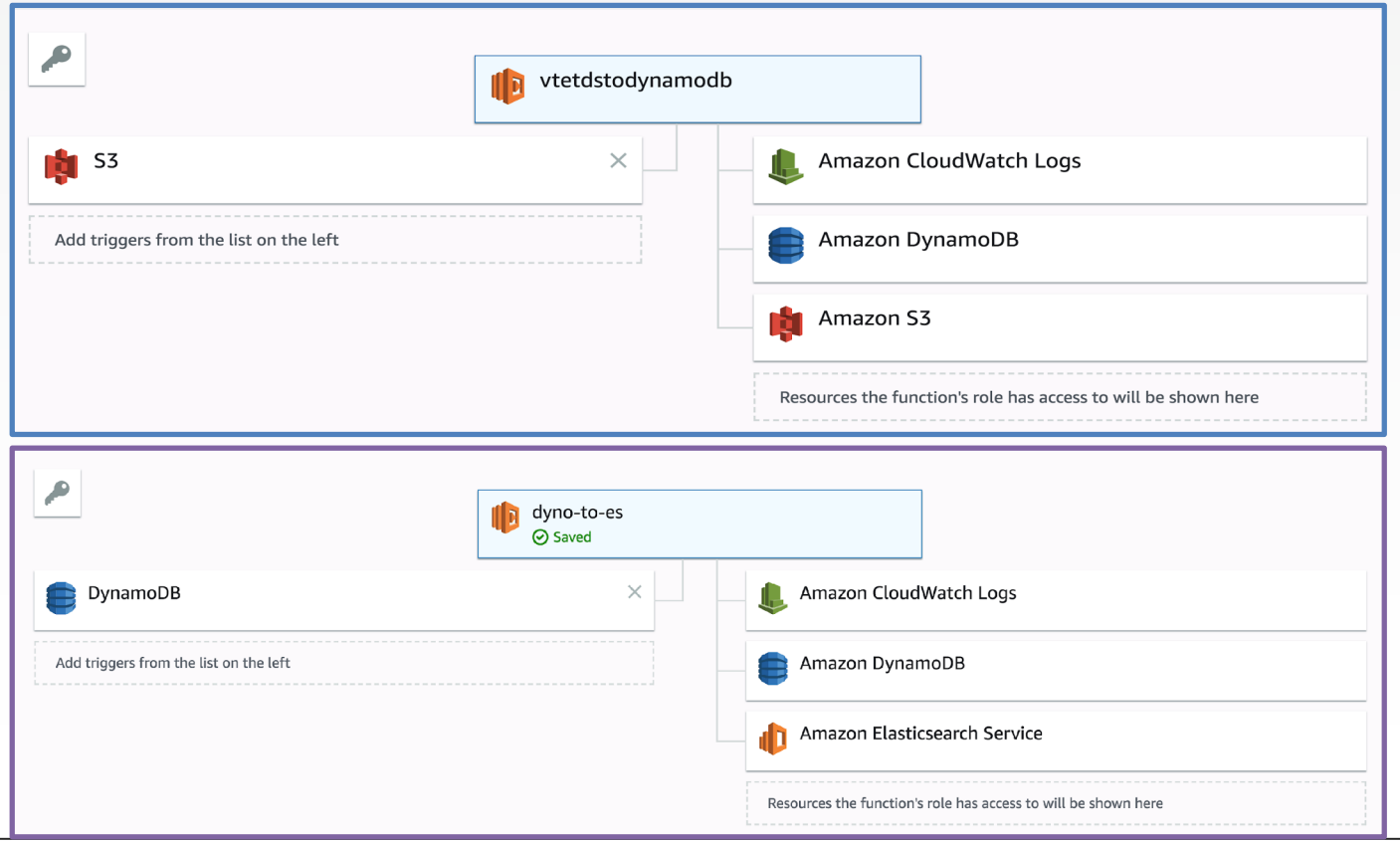
Name	Priority	SUBMITTED	PENDING	RUNNABLE	STARTING	RUNNING	FAILED
first-run-job-queue	1	0	0	0	0	0	0
IAWATileGenerationQueue	1	0	0	0	0	2	0
IAWATileGenerationQueue-c5d	10	0	0	0	0	0	0
QueueForAlexander	1	0	0	19	0	0	0
QueueForFeuerstein	1	0	0	88	0	0	0
QueueForJansone	1	0	0	0	0	50	0
QueueForRupp	1	0	0	49	0	0	0

Compute environments

Name	Type	Minimum vCPUs	Desired vCPUs	Max
i3-2XL-AMI	MANAGED	0	8	256
EnvForJansone	MANAGED	0	64	256

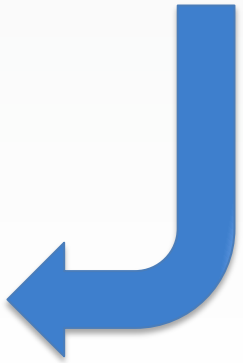
VIRGINIA TECH

Microservice – Using AWS Lambda



Metadata Transformation Using AWS Lambda

```
9 43029,10919/9291||10919/71751,,,,,"Ketene, Alperen Nurullah",,,"Agah, Masoud",,,,,,"Behkam, Bahareh||Schmelz,
  Eva M.",,Mechanical Engineering,,,,5/31/11,,,5/6/11,5/31/12,,5/18/11,,,,"According to the American Cancer
  Society, Cancer is the second most common cause of death in the United States, only exceeded by heart
  disease. Over the past decade, deciphering the complex structure of individual cells and understanding the
  symptoms of cancer disease has been a highly emphasized research area. The exact cause of Cancer and the
  genetic heterogeneity that determines the severity of the disease and its response to treatment has been a
  great challenge. Researchers from the engineering discipline have increasingly made use of recent
  technological innovations, namely the Atomic Force Microscope (AFM), to better understand cell physics and
  provide a means for cell biomechanical profiling.
```




```
{
  "author": "\"Alperen Nurullah Ketene\"",
  "collection": "10919/9291||10919/71751",
  "committeechair": "\"Masoud Agah\"",
  "degreelevel": "masters",
  "degreename": "Master of Science",
  "department": "Mechanical Engineering",
  "identifier": "etd-05182011-152552",
  "publisher": "Virginia Tech",
  "thedate": "5/6/11",
  "title": "The AFM Study of Ovarian Cell Structural Mechanics in the Progression of Cancer",
  "type": "Thesis",
  "vid": "43029"
}
```

Outcomes

- Developer/DevOps candidate pool much larger
- Automated compliance with Digital Preservation Best Practices
- Benefits of tiered storage for long-term data archiving
- Performance improvements even without optimization

Performance improvement before optimization

Page	IAWAv1	IAWAv2	Percent Change
Index Page V1 V2	796 ms	766 ms	4% faster
Item Browse Page V1 V2	1.35 s	882 ms	35% faster
Collection Browse Page V1 V2	926 ms	754 ms	20% faster
Item Page V1 V2	1.01 s	861 ms	16% faster

Site performance

Collection page

CHECKS COMPLETE **25 of 25 Locations** | ERRORS FROM **0 Locations** | AVG: 1st VISIT **8.3 sec** | AVG: 2nd VISIT **2.1 sec**

AGENT / LOCATION	FIRST VISIT	REPEAT VISIT
Minneapolis	3.4 sec	2.2 sec
New York	2.8 sec	2.0 sec
London	3.1 sec	1.5 sec
San Francisco	3.2 sec	1.4 sec
Miami	1.7 sec	1.2 sec
Hong Kong	8.3 sec	2.9 sec
Montreal	1.5 sec	1.2 sec
Frankfurt	4.0 sec	2.4 sec
Denver	2.4 sec	1.4 sec
Brisbane	11.8 sec	1.7 sec
Dallas	2.0 sec	1.4 sec
Amsterdam	4.1 sec	2.1 sec
Tel-Aviv	9.9 sec	2.1 sec
Washington DC	1.4 sec	1.1 sec
N. Virginia	2.2 sec	2.0 sec
Shanghai	31.1 sec	2.3 sec

Search page

CHECKS COMPLETE **25 of 25 Locations** | ERRORS FROM **0 Locations** | AVG: 1st VISIT **5.4 sec** | AVG: 2nd VISIT **2.0 sec**

AGENT / LOCATION	FIRST VISIT	REPEAT VISIT
Minneapolis	5.1 sec	3.0 sec
New York	2.4 sec	1.8 sec
London	2.7 sec	2.2 sec
San Francisco	2.2 sec	1.5 sec
Miami	1.4 sec	1.1 sec
Hong Kong	8.3 sec	2.1 sec
Montreal	1.3 sec	1.2 sec
Frankfurt	3.6 sec	1.3 sec
Denver	1.7 sec	1.0 sec
Brisbane	7.2 sec	1.8 sec
Dallas	1.3 sec	1.2 sec
Amsterdam	4.6 sec	1.6 sec
Tel-Aviv	4.4 sec	1.2 sec
Washington DC	1.2 sec	1.2 sec
N. Virginia	1.9 sec	1.9 sec
Shanghai	11.0 sec	1.6 sec

Demo

https://iawa-dev.cloud.lib.vt.edu/

Next Steps

- Docker and kubernetes for reproducible builds and orchestration between cloud and local
- Exploring local infrastructure changes e.g. Ceph storage
- Benchmarking and cost optimization of cloud services
- Refactoring of legacy applications to AWS
- CloudFormation or Terraform for everything

Q & A

Thank You!