

Overview of AWS infrastructure supporting the University Library's Illinois Data Bank, Digital Library and Medusa web applications

Purpose: This document provides an overview of the AWS infrastructure supporting the University Library's **Data Bank**, **Digital Library** and **Medusa** web applications. These applications provide access to, and management of, repositories of research data, digital images, and other digital scholarly content.

This document focuses on the production infrastructure, but also describes the “demo” environment used to test and showcase functionality. The audience is primarily AWS and UIUC Technical Services staff that may help Library IT support these applications.

Quick Facts:

- A Suite of web apps supporting 140 TB of digital scholarly content.
- Ruby on Rails user interfaces.
- Uses Postgres relational databases, S3 buckets, Elasticsearch and on-campus storage to manage content.
- A microservices architecture utilizing EC2 servers as well as ECS/Fargate (serverless).
- Developed and supported by Library IT.
- In addition to the “**Production**” environment, the AWS implementation also includes a “**Demo**” environment that resides in the same VPC (virtual private cloud).
- In production with AWS infrastructure since February 2019.
- Terraform (Infrastructure as Code - IaC) used to manage AWS resources. Ansible is used to configure EC2s. Capistrano is used to deploy application (Ruby) releases.

Table of contents:

1. Web Applications hosted on EC2 servers (**production**):
2. Supporting microservices hosted on EC2 & On-campus servers (**production**):
3. Supporting databases on RDS instances (**production**):
4. Supporting Elasticsearch Clusters (**production**):
5. Supporting Services on ECS / Fargate (**production**):
6. Diagram of major production AWS resources (**production**):
7. Summary of AWS resources (**production** and **demo**):
8. Demo environment components (**demo**):
9. VPC details (**production** and **demo** use the same VPC):

Appendices: (Reference and supplemental information used to create the above sections)

1. **Web Applications hosted on EC2 servers (**production**):**

- *Illinois Databank* - databank.illinois.edu
 - Publicly accessible repository for publishing UIUC research data.
- *Digital Library* - digital.library.illinois.edu
 - Illinois Library Digital Collections.
- *Medusa* - medusa.library.illinois.edu
 - Preservation management services supporting digital collections.
- *Ideals* (to be migrated to AWS in 2020) - currently ideals.illinois.edu
 - Online repository of scholarly content and research of UIUC faculty, staff and students.

2. Supporting Microservices hosted on EC2 & On-campus servers (production):

- *Copy Server* - **On-campus Virtual Machine**
 - Copies content from on-campus staging locations to AWS S3 buckets.
 - On campus VM is transfer-digitize.library.illinois.edu
- *Databank Tasks* - EC2
 - Provides specific services needed for Illinois Databank.
- *Downloader* - EC2
 - Provides content downloading functionality for applications.
- *Globus* - EC2
 - Implements Globus service (<https://www.globus.org/>) to upload and download very large datasets for Illinois Databank. Public facing subnet
- *Cantaloupe* - *iiif image server* - EC2
 - Provides image management services for applications.
 - **Note:** We also have an implementation of Cantaloupe that runs with ECS and Fargate.
- *RabbitMQ* - EC2
 - Manages messages between applications and services.
- *Solr* - EC2
 - Provides search functionality for applications.

3. Supporting databases on RDS instances (production):

- *book-tracker-prod*
 - Book_tracker postgres DB - supports book tracker service.
- *medusa-prod-rds*
 - databank postgres DB - supports databank web app.
 - databank_tasks postgres DB - supports databank task service.
 - downloader postgres DB - supports downloader service.
 - kumquat postgres DB - supports the digital library web app.
 - medusa postgres DB - supports the medusa web app.
- *metaslurp*
 - metaslurp postgres DB - supports metaslurp (metadata harvesting) service.

4. Supporting Elasticsearch Clusters (production):

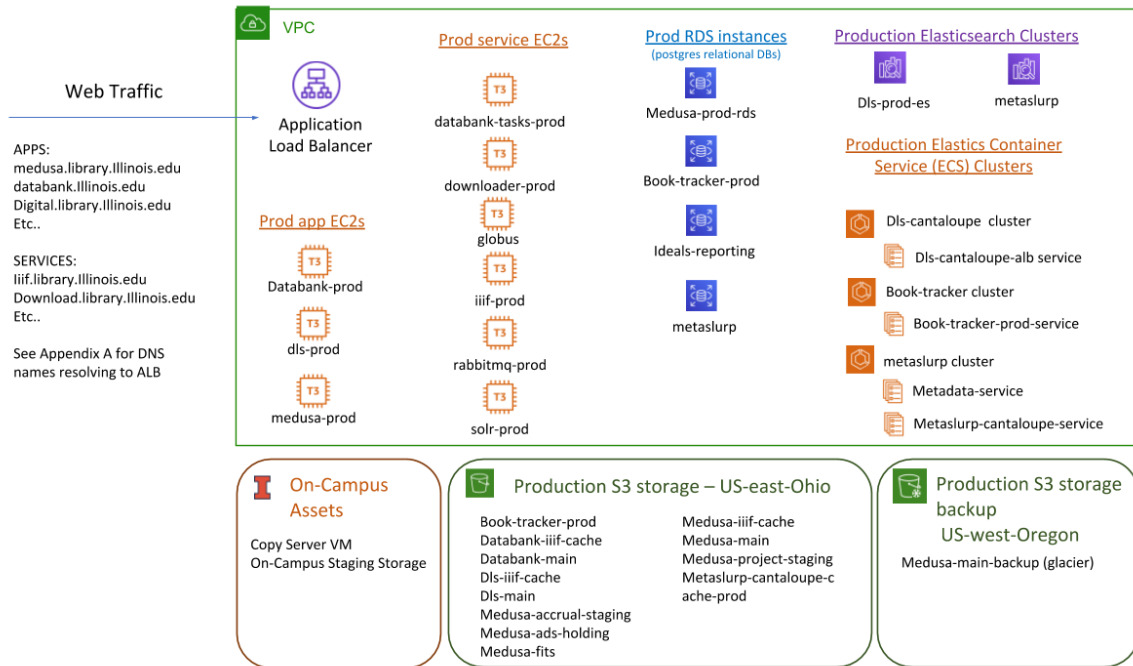
- *dls-prod-es*

- *metaslurp*

5. Supporting Services on ECS / Fargate (serverless) (production):

- Book Tracker
 - Maintains a database of digitized books (Rails / PostgreSQL App).
- Metaslurp
 - Harvest content and create searchable central index.
- Cantaloupe
 - Provides image services for the Digital Library.
- Digital Library Search Gateway - *metadata.library.illinois.edu*

6. Diagram of major AWS resources (production):



7. Summary of production AWS resources (production and demo):

AWS resource	Use
EC2, EBS	Servers to run web applications and services
ALB	Application load balancer used to route web traffic to EC2s
ECS, Fargate	Services that support the web applications
Lambda	Monitor VPN Status
S3	Storage for digital content, terraform state data, log data
S3 - CRR	Cross Region Replication to backup medusa content.
EFS	Temporary file storage for web applications
RDS	PostgreSQL Databases used by the web applications
Aurora	Used by some PostgreSQL databases in our demo environment
CloudWatch	Monitor VPN status
VPC	Virtual Private Cloud communicates with campus VPN.

8. Demo environment components (demo):

Quick “demo” facts:

- Used for Training, Testing, demonstrating new features.
- Architecture similar to production (a few differences).
- Smaller EC2 instances and EC2s are started and stopped by the “AWS Instance Scheduler”. On during business hours.
- A separate demo ALB to route internet traffic to demo servers.
- Unlike production, we use Aurora PostgreSQL for 2 of our databases.

8.1 Web Applications hosted on EC2 servers (demo):

- *Illinois Databank* - *demo.databank.illinois.edu*
 - Public Access repository for publishing UIUC research data.
- *Digital Library* - *demo.digital.library.illinois.edu*
 - Illinois Library Digital Collections.
- *Medusa* - *demo.medusa.library.illinois.edu*
 - Preservation management services supporting digital collections.
- *Ideals* - *demo.ideals.illinois.edu*
 - This EC2 is being used as a test environment for the IDEALS re-write that will migrate the application to a ruby-on-rails implementation.

8.2 Supporting Microservices hosted on EC2 & On-campus servers (demo):

- *Copy Server* - **On-campus Virtual Machine - same VM used for production.**
 - Copies content from on-campus staging locations to AWS S3 buckets.

- *transfer-digitize.library.illinois.edu*
- *Databank Tasks - EC2*
 - Provides specific services needed for Illinois Databank.
- *Downloader - EC2*
 - Provides downloading functionality for applications.
- Globus - not yet implemented in demo
- *Cantaloupe - iif image server - - EC2*
 - Provides image management services for applications.
- *RabbitMQ and SOLR - EC2 - both apps use the same EC2*
 - Manages messages between applications and services.

8.3 Supporting databases on RDS instances (demo):

- *book-tracker-demo-rds-serverless - Uses Aurora PostgreSQL*
 - Book_tracker postgres DB - supports book tracker service.
- *medusa-demo-rds*
 - databank postgres DB - supports databank web app.
 - databank_tasks postgres DB - supports databank task service.
 - downloader postgres DB - supports downloader service.
 - kumquat postgres DB - supports the digital library web app.
 - medusa postgres DB - supports the medusa web app.
- *Metaslurp-demo-rds-serverless Uses Aurora PostgreSQL*
 - metaslurp postgres DB - supports metaslurp (metadata harvesting) service.

8.4 Supporting Elasticsearch Clusters (demo):

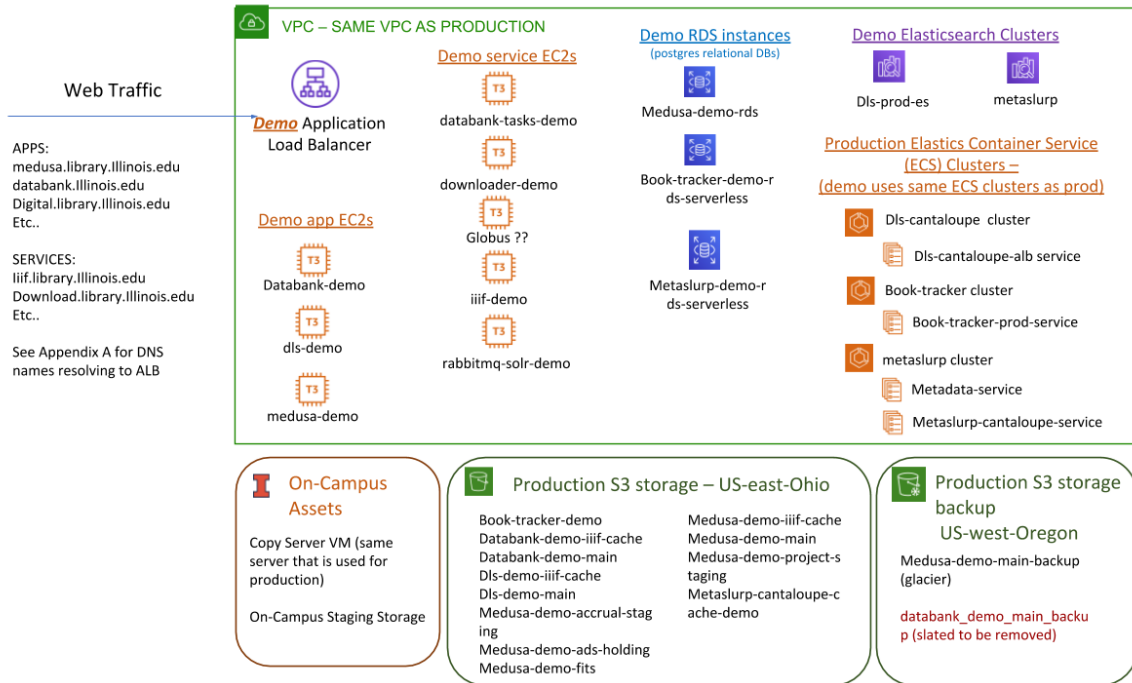
- *Dls-demo-es - One t2.small.elasticsearch in cluster*
- *Ideals-demo - One t2.small.elasticsearch in cluster*
- *Metaslurp-demo - One 1 t2.medium.elasticsearch in cluster*

8.5 Supporting Services on ECS / Fargate (serverless) (demo):

The same ECS clusters are used as in production, but there are different services and task definitions for demo.

- Book Tracker
 - Maintains a database of digitized books (Rails / Postgres App).
- Metaslurp
 - Harvest content and create searchable central index.
- Cantaloupe
 - Provides image services for the Digital Library.
- Digital Library Search Gateway - *metadata.library.illinois.edu*

8.6 Diagram of Major AWS resources (demo)

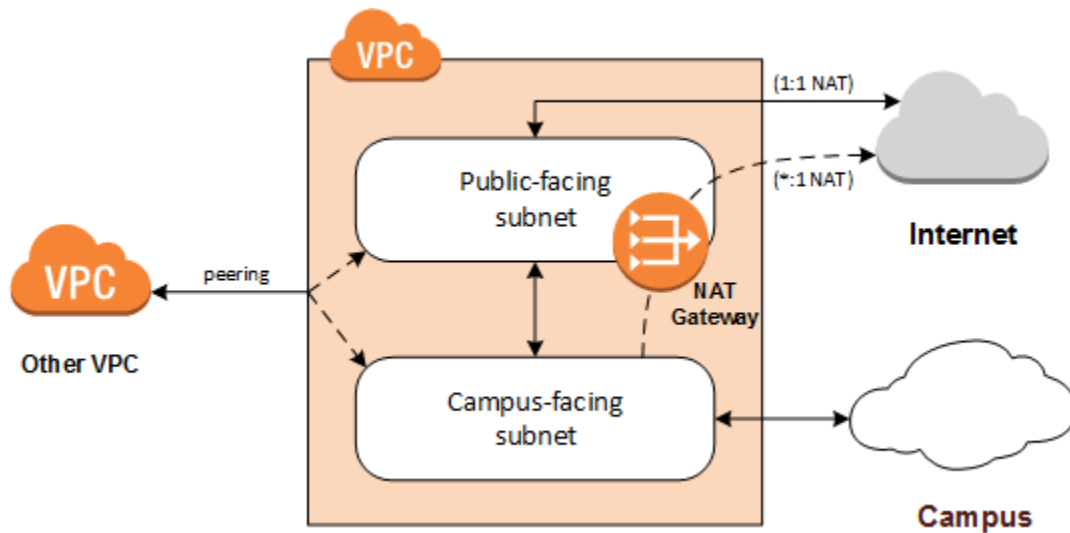


9. VPC (Virtual Private Cloud) details (production and demo use the same VPC):

Quick VPC facts:

- **Library-medusa1-vpc** uses the "VPC" template designed by UIUC Tech Services to achieve consistent security and access features for all Campus VPCs. This VPC uses OPTION 2 as described here: <https://answers.uillinois.edu/illinois/page.php?id=74081>
- **Library-medusa1-vpc** is an "UIUC Enterprise VPC" which allows campus facing subnets and peering to other "UIUC Enterprise VPCs".
- Allows resources in the VPC to communicate to resources in the campus VPN.
- More information on Tech Services Github: <https://github.com/techservicesillinois/aws-enterprise-vpc>
- The guide for using a campus VPC is here: <https://answers.uillinois.edu/illinois/page.php?id=71015>
- UIUC implements an Infoblox (infoblox.com) DNS management solution. Implemented as "IPAM" ipam.illinois.edu
- IPv4 CIDR block allocated by tech services = 10.225.250.0/24 (First IP = 10.225.250.0, Last IP = 12.225.250.255, total IPs = 256)

9.1 Simple Campus VPC diagram (source = <https://answers.uillinois.edu/illinois/page.php?id=71015>)



Appendices (Reference information used to create the above sections)

Appendix A:

IPAM CNAME records directed to the **production** load balancer :

Source: IPAM ipam.illinois.edu (using IPAM, search on DNS Name = medusa-alb-817700096.us-east-2.elb.amazonaws.com, and then select all results and export to csv file)

Name	Type	Matched Property	Matched Value (Production ALB DNS Name)
default/medusa-alb.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/aws-download-alb.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/aws-iiif-alb.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/aws-databank-alb.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/aws-medusa-alb.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/aws-dls-alb.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/aws-dls-cantaloupe-alb.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/aws-metadata-alb.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/metadata-dev-fargate.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/metadata.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/medusa.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/images.digital.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/digital.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/download.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/iiif.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com

default/databank.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/www.databank.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/metdata-dev-fargate.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/medusahq.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/medusausers.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/demo.images.databank.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/book-tracker.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com
default/images.metadata.library.illinois.edu	CNAME Record	Canonical Name	medusa-alb-817700096.us-east-2.elb.amazonaws.com

IPAM CNAME records directed to the **demo** load balancer:

Source: IPAM - ipam.illinois.edu (using IPAM, search on DNS Name = [internal-demo-medusa-alb-172688561.us-east-2.elb.amazonaws.com](#), then select all results and export to csv file)

Name	Type	Matched Property	Matched Value (Demo ALB DNS name)
default/demo.databank.illinois.edu	CNAME Record	Canonical Name	internal-demo-medusa-alb-172688561.us-east-2.elb.amazonaws.com
default/demo.medusahq.library.illinois.edu	CNAME Record	Canonical Name	internal-demo-medusa-alb-172688561.us-east-2.elb.amazonaws.com
default/demo.medusausers.library.illinois.edu	CNAME Record	Canonical Name	internal-demo-medusa-alb-172688561.us-east-2.elb.amazonaws.com
default/demo.medusa-alb.library.illinois.edu	CNAME Record	Canonical Name	internal-demo-medusa-alb-172688561.us-east-2.elb.amazonaws.com

default/demo.book-tracker.library.illinois.edu	CNAME Record	Canonical Name	internal-demo-medusa-alb-172688561.us-east-2.elb.amazonaws.com
default/rds-dev.library.illinois.edu	CNAME Record	Canonical Name	internal-demo-medusa-alb-172688561.us-east-2.elb.amazonaws.com
default/medusadev1.library.illinois.edu	CNAME Record	Canonical Name	internal-demo-medusa-alb-172688561.us-east-2.elb.amazonaws.com
default/medusatest.library.illinois.edu	CNAME Record	Canonical Name	internal-demo-medusa-alb-172688561.us-east-2.elb.amazonaws.com
default/kumquat-dev.library.illinois.edu	CNAME Record	Canonical Name	internal-demo-medusa-alb-172688561.us-east-2.elb.amazonaws.com
default/demo.ideals.illinois.edu	CNAME Record	Canonical Name	internal-demo-medusa-alb-172688561.us-east-2.elb.amazonaws.com
default/demo.metadata.library.illinois.edu	CNAME Record	Canonical Name	internal-demo-medusa-alb-172688561.us-east-2.elb.amazonaws.com
default/demo.images.metadata.library.illinois.edu	CNAME Record	Canonical Name	internal-demo-medusa-alb-172688561.us-east-2.elb.amazonaws.com
default/demo.iiif.library.illinois.edu	CNAME Record	Canonical Name	internal-demo-medusa-alb-172688561.us-east-2.elb.amazonaws.com
default/aws-download-demo.library.illinois.edu	CNAME Record	Canonical Name	internal-demo-medusa-alb-172688561.us-east-2.elb.amazonaws.com
default/demo.medusa.library.illinois.edu	CNAME Record	Canonical Name	internal-demo-medusa-alb-172688561.us-east-2.elb.amazonaws.com
default/demo.digital.library.illinois.edu	CNAME Record	Canonical Name	internal-demo-medusa-alb-172688561.us-east-2.elb.amazonaws.com

Appendix B:

Production App Servers (EC2s):

databank-prod t3.large
dls-prod t3.xlarge
medusa-prod t3.xlarge

--

Demo App Servers (EC2s):

databank-demo	t3.medium
dls-demo	t3.xlarge
Ideals-demo	t3.medium
medusa-demo	t3.large

Appendix C:

Production Service Servers (EC2s):

Databank-tasks-prod	t3.xlarge
Downloader-prod	t3.medium
Globus-pilot	t3.small
liif-prod	t3.xlarge
Rabbitmq-prod	t3.medium
Solr-prod	t3.2xlarge

Transfer-digitize.library.illinois.edu - (copy server) ON-Campus VM - both prod and demo

--

Demo Service Servers (EC2s):

Databank-tasks-demo	t3.small
Downloader-demo	t3.small
liif-demo	t3.small
Rabbitmq-solr-demo	t3.large (for demo both rabbitmq and solr live on the same server)

Transfer-digitize.library.illinois.edu - (copy server) ON-Campus VM - both prod and demo