Lowering the Barrier to Entry for Digital Repository Management by Leveraging Cloud-Native Solutions

Open Repositories

June 4th, 2024

Favenzio Calvo & Bryan J. Brown

Florida State University Libraries

About Us

Favenzio Calvo

- Director of Software Development
- Joined FSU Libraries in 2012

Bryan J. Brown

- Repository Developer
- Joined FSU Libraries in 2014

About Our Organization

Florida State University

We are a public research university located in the United States in Tallahassee, Florida. We serve a population of approximately 43,000 students and 2,600 traditional faculty. The FSU Libraries have been charged with preserving many of the scholarly outputs of our community, as well as digitizing a vast number of rare and unique cultural heritage items that reside in our Special Collections unit.

Presentation Goals

Introduce

Introduce a few key concepts related to cloud migrations and digital repositories

Share

Share our challenges and lessons learned from migrating to the cloud

Discuss

Provide some ideas and discussion topics related to the state of open repositories today, and their relationship to cloud services

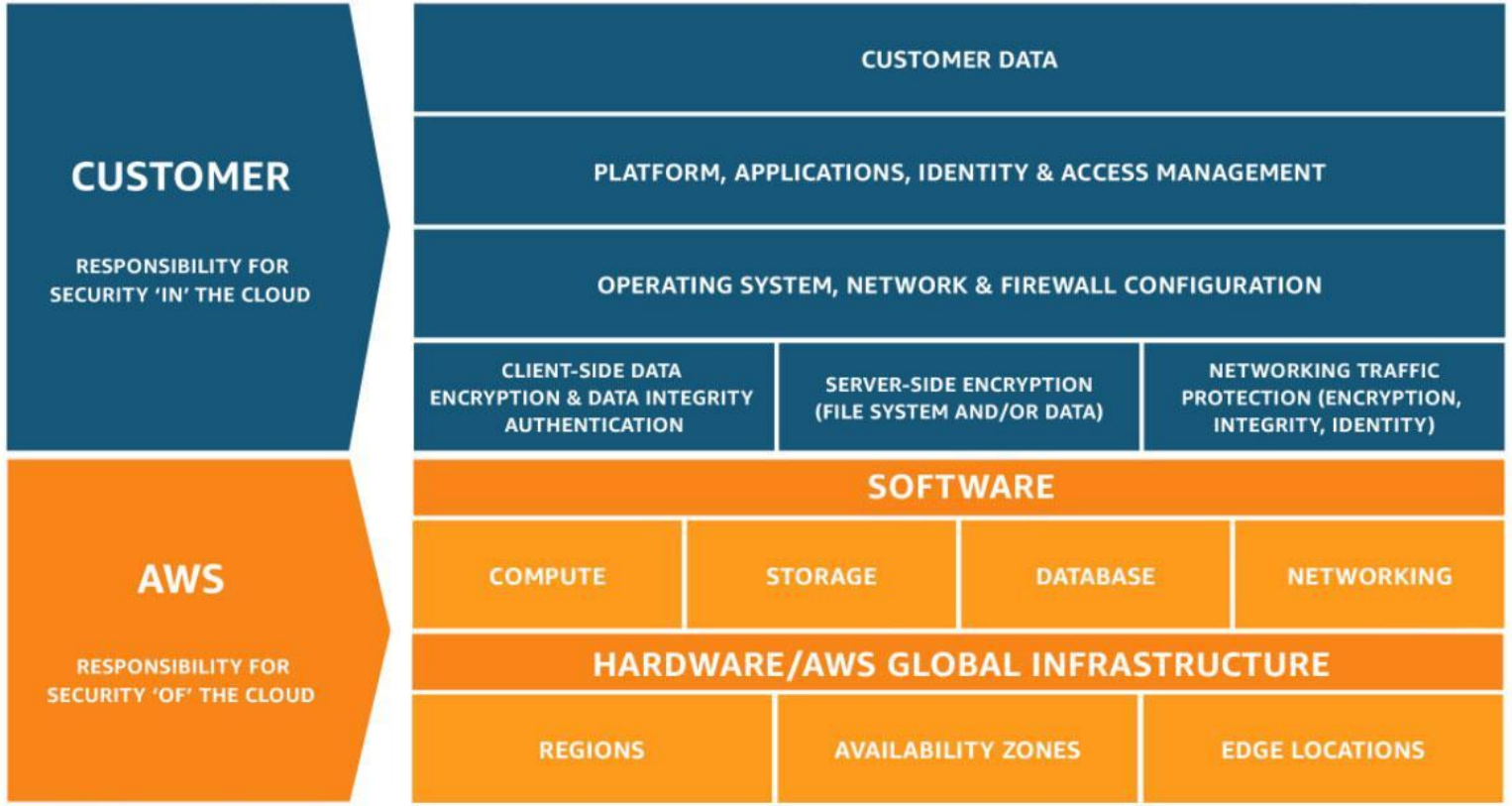
!

DISCLAIMER

"We could not, with our own staff that covers data, coding, documentation, research and support, run the operation responsibility right now without (Cloud Provider)."

Diego Pino, Director of Digital Strategy
Metropolitan New York Library Council
Slack Conversation, May 1, 2024

AWS Shared Responsibility Model



Retrieved from https://aws.amazon.com/compliance/shared-responsibility-model/ on May 20th, 2024

Cloud Enabled vs Cloud Native

Cloud Enabled

- ✔ Originally designed monolithically with some tightly integrated components, but adapted to work in the cloud
- ✔ Usually have limited scalability or capabilities to dynamically allocate resources
- ✔ May incur higher costs due to lack of optimized resource utilization

Cloud Native

- ✔ Built from the ground up to leverage cloud computing to its fullest extent
- ✔ Typically follow a distributed microservices architecture, allowing for scalability and fault tolerance
- ✔ Resource usage optimization can achieve higher cost savings

Our Journey into the Cloud

Pre-cloud Infrastructure (2015)

- Applications
  - Library managed in-house Dell RHEL servers
  - Campus IT Services managed virtual servers in a local data center

- Storage
  - Library managed in-house NAS servers
  - Campus IT Services managed storage in a local data center

Our Journey into the Cloud

Pre-cloud Challenges (2015)

- Provisioning
  - Requires up-front estimation
  - Moving at the speed of campus bureaucracy

- Availability & Continuity of Operations
  - Lack of distributed storage
  - Multiple single points of failure

Our Journey into the Cloud

The Procurement Process

Challenges

- Variable Costs
- Data Rights and Privacy Concerns

Solutions

- Spending Limit Contract & Cost Alarm Thresholds
- In-House Counsel Contract Review

Our Journey into the Cloud

Transitioning to AWS (2016)

- Started with a lift-and-shift of web services
  - Recreate as similarly as possible, but in new environment
  - Least important/complicated services first
- Gradual optimizations
  - Including more managed services
    - Managed services are low maintenance & high reliability. Often cheaper too.
    - MVPs: RDS (database) & S3 (storage)
  - Redesigning deployments
    - Version-controlled deployment templates
    - Elastic Load Balancers & autoscaling + health checks
    - Constantly re-evaluating resource needs

Our Journey into the Cloud

Digital Repository Migration to AWS (2021)

- DigiNole (Islandora 7) repository: most complicated, last to be migrated
- Our first production containerized application (ISLE customization)
- Cloud repository storage can be the most expensive part
  - Can your repository platform make use of object storage (like S3)? If not, $$$
  - Does your repository platform store derivatives (duplicate copies of assets in different formats)? If so, $$$+$$$+$$$
  - Does your repository platform's storage facilitate easy backup & restore? If not, ($$$+$$$+$$$) x ($$$+$$$+$$$)
- Not always easy/possible to break up monolithic apps (Cloud Enabling has limits)
  - Localhost vs private networks
  - Repeated micro-lag becomes macro-lag

Our Journey into the Cloud

Where Are We Going?

- Experience changes how we evaluate software & services
  - Storage-heavy services NEED to be compatible with managed object storage
  - Search-heavy services SHOULD be compatible with managed search solutions
  - Database-driven services NEED to be compatible with managed database solutions
  - Non-dynamic applications SHOULD be replaced with static sites when possible
- Experience changes how we plan new deployments
  - KISS: Keep It Simple, Sysadmins
  - Automate as much as you can
  - Modernize legacy deployments

How Can Cloud Services Support Repositories?

- Managed services
  - Servers
  - Databases
  - Storage
- Inclusivity
  - Lower barrier to entry
  - Democratized access
- Sustainability
  - Energy
  - Materials
  - Space

What Are the Risks with Cloud Services?

- Unchecked usage charging
  - Storage
  - Autoscaling
  - Forgetting to turn off services
- Security
  - Identity management
  - Firewalls
- Regulatory concerns
  - Is your data secure?
  - Where does the data live?
  - Can you prove it hasn't been altered?

Summary

- ✔ The cloud offers benefits, but also challenges
- ✔ Migrate your services, staff and culture
- ✔ Start with a lift-and-shift
- ✔ You're never done optimizing
- ✔ For repositories, the biggest area for cost savings is storage

Current State of Cloud Repositories?

- Are modern Open Source repositories Cloud Enabled / Cloud Native?
- Does the complexity and large number of cloud providers / solutions make it too difficult for Open Source communities to provide more support?
- To what degree are Open Source communities thinking about the Cloud and providing resources for smaller to medium size institutions? What more can we do?

Questions or Comments?