Cloud Migration: Issues and Solutions

Group 5: Nathan Lederach, Emanuel Lazar, Ruchita Paithankar

INFO 540

Drexel University

September 2, 2022

**Summary**

EmDeployment is migrating a second information system (IS) to the Amazon Web Services (AWS) cloud environment.   The company’s content management system (CMS) is having latency and security issues at its own premise data center and hopes to take advantage of the AWS content delivery network (CDN) to improve performance and security.  The company also wants to address poor turnaround time for video encoding and is looking to adopt an automated serverless workflow process that increases time-to-market availability for high-resolution video content. Lastly, because this is the second IS the company will integrate into its AWS account, a plan is presented that addresses how a cloud migration strategy will be consistently applied for subsequent migrations.

**Information System Plan**

Information System Planning helps realize a roadmap to meet objectives and define metrics of success (Mirchandani & Lederer, 2012).  EmDeployment’s mission is to become a global leader in career and job counseling, providing tailored services to different customer segments depending on their country of residence. To realize this goal, we will develop a suite of performant high availability applications.

Customer-facing applications will be migrated to a cloud service provider first. Internal systems used by EmDeployment staff like customer relationship management (CRM) and Human Resources (HR) systems do not require the same high availability and performance as external customers and will not be part of the first systems to migrate.  By focusing on externally facing systems, we can increase customer subscriptions and customer engagement.

**Analysis of Current Problems**

Our on-premises data center no longer meets our Information Technology (IT) needs.  Our current infrastructure is in a single geographic location. This has exacerbated the effects of extreme weather events and distributed denial of service (DDoS) attacks that in aggregate have made the availability of our systems only 99%, costing us an average of four days of downtime per year (Lerena, 2022). In addition, our global business growth is suffering because internet traffic to countries like India or Korea is prone to high latency of more than 300 milliseconds due to the geographic distance between our U.S. data center and customer endpoints in Asia (Ali et al., 2017). This is causing customer complaints and loss of business. Lastly, Global trends in regulatory requirements around data localization and data handling have become onerous in terms of compliance, forcing us to expend resources on the legality of operations rather than business growth. Cloud hosting can solve these issues if we deploy consistent system architectures with those baseline requirements in mind.

**Greenfield Migration**

  The migration method of IT systems will use a greenfield approach (Sharma, 2020). By redesigning legacy applications in a cloud environment, we abandon any dependencies on our existing outdated technology and start fresh using the scalable, innovative, and optimized technologies available in the cloud. Although this will require new policies and practices as well as reskilling staff or hiring new cloud solutions architects, it will allow us to create applications with our technology goals of high availability and increased performance in mind instead of lifting and shifting our current IT problems to a new environment.

**Reference Architecture**

  Every system in the cloud will use a reference architecture for high availability (HA) that incorporates multi–Availability Zone (AZ) or multi-region failover that provides at least 99.95% availability (Filecloud, 2022).  Default architecture will consider low latency as a business requirement and only use cloud services that prioritize low latency routing to increase performance (Chauhan et al., 2018, p. 473). It will also be a baseline requirement that cloud services have regulatory compliance attestation for the markets in which we want to do business. Lastly, we will use cloud providers that limit the geographic distribution of data in accordance with data localization requirements. We have taken care of this in the previous EmDeployment migration by hosting customer-focused applications that collect personally identifiable information (PII) in Europe to comply with General Data Protection Regulation (GDPR), but as more applications are added to the cloud it will be imperative to keep up with global regulatory requirements. The cloud provider will absorb some of that burden because we will take advantage of a shared responsibility model for security and compliance (Wexler, 2022).

**Organizational Unit**s

As more applications migrate to the cloud, we can track the cost and performance of each application using OU’s.  OU’s will logically segment projects and provide transparent billing, which will justify the IT cost of cloud transition and operation. OU’s divide the billing into unique categories to better define what is the cost in each information system. EmDeployment will have two types of accounts: master and child. The child accounts will host all the resources from each EmDeployment IS, while the master accounts will be used to collect, monitor, and report costs. We will also use Cost Explorer in AWS Cost Management Center, as this will help visualize cost. We can also see data transfer charges from region to region. Additionally, tags will be used to keep track of costs at a detailed level. This will help us to organize how much resources cost while making it easier to track our expenses (Using Cost Allocation Tags - AWS Billing, 2022).

**Content Delivery**

**Overview of Problem**

EmDeployment provides a website that hosts a variety of training materials delivered as static content (i.e. videos and images) on servers in its on-premises data center in the United States (US). These training videos are marketed to potential job seekers all around the globe who are looking for career opportunities in the US or Europe. They explain useful ex-pat needs like how to secure a visa and how to relocate overseas for work. The videos are high quality and can be lengthy, which is causing poor performance of the website. There are several issues with the existing solution. We are finding that for customers in Asia, websites can delay loading, which is causing customer disengagement. There are also issues around costs and time associated with converting videos that are acceptable formats for viewing over the internet. The U.S.-based data center is susceptible to DDoS attacks because it is a single point of failure, and our infrastructure is not equipped to deal with sophisticated DDoS attacks.

**Recommendation**

The recommendation is to migrate the TrainVid application to the global AWS CDN CloudFront. This will allow us to reduce web page load time, reduce bandwidth cost, increase content availability, and provide additional security against DDoS attacks.  Data can be cached at edge locations around the globe, moving data closer to the customers and reducing latency (Shannon, 2018). AWS elemental media services will be used to take source content and convert it before storage in a storage account, automating this cumbersome workflow process with Lambda functions to reduce cost and increase efficiency (Rapyder, 2022). The CDN solution will also automatically redistribute traffic when needed to provide DDoS attack prevention through the AWS shield standard, thereby increasing the uptime of the site.

**Latency Problems**

The TrainVid content management system (CMS) hosts a variety of videos in MP4 format on a static website and provides the content to customers from around the globe.   Users access our paid content through servers at our on-premises data center in Austin, TX.   The company has recently had an increase in customers in Asia, and new customers in that region are commenting on how the website often buffers and slows down.  Some users have canceled subscriptions because of this.  The poor performance of the website for our customers in Asia can be attributed to the distance between the client device and our data center in the United States.  This distance is causing high latency, which makes web pages load slowly (Gillis, 2020).   Looking at average latency statistics, we discovered North America to India latency to average 230 milliseconds and North America to Korea to be 150 milliseconds (Verizon, n.d.).  We are looking to achieve latency rates of under 100 milliseconds as a metric of success (Fitzgerald, 2022).

**Cloud Solutions**

AWS CloudFront CDN solves the latency issue by bringing data closer to users around the world.  This is done using edge locations, which are regional points of presence (PoP) to locally cache data.  Amazon has hundreds of edge locations around the globe, including in areas like Asia where we hope to increase business performance (Chauhan et al., 2018).  In this solution, we would host our static video files in an origin server which would be an Amazon Storage (S3) bucket (Chauhan et al., 2018).  As users make requests for the files, latency-based routing steers them to the closest edge location.  Once the data is loaded from the origin server, it remains cached in the Edge Location.  This allows users in one geographic area to have low latency access to the data for as long as it remains cached. The data remains cached for a specified period, which could be months because our videos are static content that do not need frequent version updates  (A.W.S., n.d.).

**Independent Testing of CDN Performance**

CloudFront provides built-in performance monitoring and metrics that are beneficial to the business operations, including which objects have been accessed and from what edge location(A.W.S., n.d.-b).  This will help us determine the geographic areas that have the most traffic and will help steer business decisions toward those markets  However, to see if the CloudFront solution is performing within our metrics of success there is a need to deploy last mile testing.  In this case, we would need to independently verify if the geographic areas we intend to reach with low latency traffic are seeing increased performance due to the Edge Location deployments.  We will contract with a third-party performance testing company based in that geographic area and task them with observing web traffic from our site.  We will ensure the third party is using a common internet service provider (ISP) for the region and is on a computing infrastructure that would be common for our end users.  This way, we can test how our content is being delivered to actual users around the globe (Rayburn, n.d.).

**Organizational Units**

**Security**

The CMS application migration will incorporate our IS plan for cloud application segmentation because TrainVid will be its own OU in the company’s AWS account. An OU in AWS is a container for accounts. We will use OU’s to segment the security so that permissions are separate. In addition, we will also use OU’s to segment billing so we can see the cost. Security will be segmented by setting up an Identity and Access Management (IAM) account. The three main parts that make up IAM are Who, Can Access, and What (A Least-Privilege Journey: AWS IAM Policies and Access Analyzer (55:59), 2022). The Who section deals with any human access that is needed. The Can Access deals with permissions on a system. The last section deals with what users can access. We will begin to manage our AWS environment by making multiple accounts to separate the workload across multiple environments. We will govern the accounts by using AWS Organizations because it will allow us to centrally manage our environment (AWS Organizations - Amazon Web Services, 2022).  Next, we will use AWS IAM Identity Center where we can connect our users and manage access to AWS accounts and applications. Once our users are connected, we will organize our resources with tags. Tags, in AWS, will help us manage, identify, and organize our resources. By setting up data perimeters in AWS, it will help set boundaries and protect data across accounts like trusted identities and resources. Permissions are set to grant access to users. This is done by setting up IAM policies to users, groups, or roles, and resource policies that govern resources such as IAM Roles

**DDoS Protection with Amazon Shield**

Our IS plan is to have high availability and that would need us to protect our site from attacks like DDoS to reduce downtime. DDoS protection solutions that are installed on-premises at the company’s data center are not enough to protect against the huge volume of DDoS attacks which flood the internet pipe and are enough to render downstream (Chinnasamy, 2020). Plus, they are ineffective in protecting our website which is hosted on AWS. AWS Shield, on the other hand, would help us safeguard our applications that were migrated to AWS, with its always-on detection and automatic inline mitigation would minimize downtime and latency of our website and it is offered in every AWS Region (AWS, n.d.-c).

Since our company has already migrated to AWS, we have the automatic protection of AWS Shield Standard at no additional cost, but it only defends against the most common and frequently occurring transport and network layers DDoS attacks that target our website (AWS, n.d.-c). Most of our applications are currently running on Amazon Elastic Compute Cloud (EC2), Elastic Load Balancing (ELB), and Amazon Route 53; with our second IS migration incorporating CloudFront, we require a higher level of protection against attacks. The recommendation is to subscribe to AWS Shield Advanced which would provide us with additional detection and mitigation against sophisticated DDoS attacks along with the basic network and transport layer protections. AWS Shield Advanced, available globally on all the edge locations, gives near real-time visibility into attacks by integrating to AWS web application firewall (WAF) and shielding against DDoS-related spikes in our EC2, ELB, CloudFront, and Route 53 charges (Ferroni, 2022).

We would have 24x7 access to AWS Shield Response Team (SRT) which would work on the company’s behalf to sort the incidents, analyze the causes, and apply mitigations (AWS, n.d.-c). The DDoS cost protection will work against scaling charges which usually result from DDoS-related usage spiked on AWS protected resources. If any of the resources scale up as a response to a DDoS attack, Shield Advanced service credit can be requested through the regular AWS Support channel (AWS, 2020). We can also define a health check and various traffic attributes in Route 53 and associate it with resources protected with Shield Advanced through the console or API to improve responsiveness, and accuracy in attack detection and alleviation, and prevent false positive notifications (AWS, 2020). This health status would be available to the SRT.

**AWS Elemental MediaConvert**

**Serverless Workflow**

As part of our greenfield approach to cloud migration, we want to retire our on-premises video encoder and incorporate the AWS cloud service, AWS Elemental MediaConvert as part of a serverless automated workflow.   
 Conversion of videos, so that they can be seen over a broad range of sizes, formats, and bitrates, is difficult because we have them in huge numbers (AWS, n.d.-d). We already have an extensive video repository stored in S3 which can be easily transcoded according to the requirements. AWS Lambda functions are evoked for these existing input videos in an S3 source bucket which then calls the Elemental MediaConvert to perform large-scale video translation and store it in an S3 destination bucket (AWS, n.d.-d). Elemental MediaConvert will help us with easy and reliable transcoding of the on-demand content for broadcast and multiscreen delivery on a global scale, with full control over the video quality and pay-as-you-go billing service (AWS, n.d.-e). We will no longer need to maintain a license for our media encoding portion of the content management system.

**References**

*A least-privilege journey: AWS IAM policies and Access Analyzer (55:59)*. (2022). Amazon Web Services, Inc. <https://aws.amazon.com/iam/getting-started/>

Ali, Wang, G., Cottrell, R. L., & Masood, S. (2017). Internet Performance Analysis of South Asian Countries Using End-to-End Internet Performance Measurements. 2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC), 2017, 1319–1326. https://doi.org/10.1109/ISPA/IUCC.2017.00201

A.K., M. (2018, September 26). *Revista ESPACIOS | Vol. 39 (No 39) Año 2018*. es.revistaespacios.com. <http://es.revistaespacios.com/a18v39n39/18393922.html>

A.W.S. (n.d.-a). *Increasing the proportion of requests that are served directly from the CloudFront caches (cache hit ratio) - Amazon CloudFront*. Amazon Web Services, Inc. <https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/cache-hit-ratio.html>

A.W.S. (n.d.-b). *Viewing CloudFront and Lambda@Edge metrics - Amazon CloudFront*. Amazon Web Services, Inc. <https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/viewing-cloudfront-metrics.html>

AWS (2020, February 14), AWS Shield Advanced now supports health based detection, <https://aws.amazon.com/about-aws/whats-new/2020/02/aws-shield-advanced-now-supports-health-based-detection/>

AWS (n.d.-c), AWS Shield features, products, security, identity & compliance, AWS Shield, <https://aws.amazon.com/shield/features/>

# AWS (n.d. - d), *Tutorial: Batch-transcoding videos with S3 Batch Operations, AWS Lambda, and AWS Elemental MediaConvert, Amazon simple storage service (S3), user guide*, <https://docs.aws.amazon.com/AmazonS3/latest/userguide/tutorial-s3-batchops-lambda-mediaconvert-video.html>

# AWS (n.d.-e), *AWS Elemental MediaConvert, features,* <https://aws.amazon.com/mediaconvert/features/>

*AWS Organizations - Amazon Web Services*. (2022). Amazon Web Services, Inc. https://aws.amazon.com/organizations/

*AWS Organizations - Amazon Web Services*. (2022). Amazon Web Services, Inc. <https://aws.amazon.com/organizations/>

AWS presents Amazon Sumerian, AWS Elemental Media Services at AWS re:Invent. (2017). *Pivotal Sources*.

Chauhan, S., Halachmi, A., Devine, J., Lehwess, M., Matthews, N., Morad, S., & Seymour, S. (2018). *AWS Certified Advanced Networking Official Study Guide: Specialty Exam* (1st ed.). Sybex.

Chinnasamy, V. (2020, November 23), Understanding Cloud DDoS Attacks and Cloud-based DDoS Protection, Indusface, <https://www.indusface.com/blog/understanding-cloud-ddos-attacks-and-cloud-based-ddos-protection/>

ClickIT. [https://www.clickittech.com/aws/aws-organizations-best-practices/](https://www.clickittech.com/aws/aws-organizations-best-practices/ )

*Cloud Object Storage – Amazon S3 – Amazon Web Services*. (2022). Amazon Web Services, Inc. <https://aws.amazon.com/s3/>

Ferroni, R. (2022, April 13), AWS Best Practices for DDoS Resiliency, AWS Whitepaper, <https://docs.aws.amazon.com/whitepapers/latest/aws-best-practices-ddos-resiliency/contributors.html>

Filecloud. (2022, March 12). *High Availability Architecture and Best Practices | FileCloud*. https://www.filecloud.com/blog/an-introduction-to-high-availability-architecture/

Fitzgerald, A. (2022, July 12). *What Is Latency & How Do You Improve It?* HubSpot. https://blog.hubspot.com/website/what-is-latency

Gillis, A. S. (2020, January 29). *Latency*. TechTarget. <https://www.techtarget.com/whatis/definition/latency>

Guy, D. (2022, April 5). *Multi-account AWS Organizations best practices for Financial Services*.

*How to Manage AWS Costs*. (2022). Amazon Web Services, Inc. https://aws.amazon.com/startups/start-building/how-to-manage-aws-costs/

*How to Manage AWS Costs*. (2022). Amazon Web Services, Inc. <https://aws.amazon.com/startups/start-building/how-to-manage-aws-costs/>

Lerena, S. (2022, August 12). *Uptime/SLA calculator: what is an SLA and how to calculate it?* Pandora FMS Monitoring. https://pandorafms.com/blog/what-is-an-sla/

Mirchandani, & Lederer, A. L. (2012). “Less is More:” Information Systems Planning in an Uncertain Environment. *Information Systems Management*, *29*(1), 13–25. <https://doi.org/10.1080/10580530.2012.634293>

Rapyder. (2022, June 27). *8 Key AWS Elemental Media Services You Should Know In 2022*. Rapyder Cloud Solutions. https://www.rapyder.com/blogs/8-key-aws-elemental-media-services-in-2022/

Rayburn, D. (n.d.). *Comparing CDN Performance: Amazon CloudFront’s Last Mile Testing Results*. Amazon Web Services, Inc. https://media.amazonwebservices.com/FS\_WP\_AWS\_CDN\_CloudFront.pdf

Shannon. (2018). *Compute Services : Amazon CloudFront*. Skillsoft Ireland Limited.

Sharma, S. (2020, September 26). *Cloud Migration for Businesses Can Be a Challenge, Greenfield Migration Explained*. Cloud Management Insider. <https://www.cloudmanagementinsider.com/cloud-migration-for-businesses-can-be-a-challenge-greenfield-migration-explained/>

*Using Cost Allocation Tags - AWS Billing*. (2022). Amazon.Com. https://docs.aws.amazon.com/awsaccountbilling/latest/aboutv2/cost-alloc-tags.html

*Using Cost Allocation Tags - AWS Billing*. (2022). Amazon.Com. <https://docs.aws.amazon.com/awsaccountbilling/latest/aboutv2/cost-alloc-tags.html>

Verizon. (n.d.). *Monthly IP Latency Data*. Verizon Enterprise. https://www.verizon.com/business/terms/latency/

Wexler, J. (2022, January 21). *Compliance in the cloud*. CIO. <https://www.cio.com/article/189596/compliance-in-the-cloud.html>