


Case Report

Medicare meets the cloud: the development of a secure platform for the storage and analysis of claims data

Roy L. Simpson, DNP¹, Joseph A. Lee, MSEE², Yin Li, PhD¹, Yu Jin Kang, PhD³, Circe Tsui, MS⁴, Jeannie P. Cimiotti , PhD^{*,1}

¹Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta, GA 30322, United States, ²Harvard University, Boston, MA 02138, United States, ³Byrdine F. Lewis College of Nursing and Health Professions, Georgia State University, Atlanta, GA 30303, United States, ⁴Office of Information Technology, Emory University, Atlanta, GA 30322, United States

*Corresponding author: Jeannie P. Cimiotti, PhD, Nell Hodgson Woodruff School of Nursing, Emory University, 1520 Clifton Road, Atlanta, GA 30322, United States (Jeannie.p.cimiotti@emory.edu)

Abstract

Introduction: Cloud-based solutions are a modern-day necessity for data intense computing. This case report describes in detail the development and implementation of Amazon Web Services (AWS) at Emory—a secure, reliable, and scalable platform to store and analyze identifiable research data from the Centers for Medicare and Medicaid Services (CMS).

Materials and Methods: Interdisciplinary teams from CMS, MBL Technologies, and Emory University collaborated to ensure compliance with CMS policy that consolidates laws, regulations, and other drivers of information security and privacy.

Results: A dedicated team of individuals ensured successful transition from a physical storage server to a cloud-based environment. This included implementing access controls, vulnerability scanning, and audit logs that are reviewed regularly with a remediation plan. User adaptation required specific training to overcome the challenges of cloud computing.

Conclusion: Challenges created opportunities for lessons learned through the creation of an end-product accepted by CMS and shared across disciplines university-wide.

Lay Summary

Data-intensive computing has been traditionally performed on large physical servers that store data and software programs. More recently, many organizations have elected to move their data from physical servers to cloud-based platforms. Cloud-based platforms are similar to internet-based computer hardware and software platforms that can run large numbers of computers and programs at any given time. There are many cloud-based platforms available, but the most common are Amazon Web Services (AWS), Google Cloud Platform, and Microsoft Azure. In this report, we describe the process involved in transition from a physical server to the AWS cloud-based computing platform. This transition was through agreement with the Centers for Medicare and Medicaid Services (CMS) in our effort to examine the healthcare outcomes of hospitalized Medicare beneficiaries. Through a series of meetings with CMS and their technology affiliate, we developed a platform within AWS that met all the security features required by the US Department of Health and Human Services. Documents were provided that outlined criteria that were needed to secure Medicare claims; detailed documents that required signatures from data stewards at Emory University. Unexpected challenges were the learning curve associated with the use of AWS and the costs associated with a cloud-based platform.

Key words: AWS; cloud computing; CMS; technical architecture.

Introduction

Cloud computing is a modern approach to data storage and management, and data-intensive computations that eliminates the need for substantial physical media.^{1,2} It provides a secure environment through user authentication, digital transformation, and real-time monitoring of large datasets.^{3,4} The use of cloud computing frees physical storage space, reduces demand for internal information technology (IT) resources, and reduces computational costs.^{5–8} Amazon Web Service (AWS) is a third-party platform, similar to other cloud-based platforms, offers cloud-based services to decrease application downtime. This is accomplished through the reduction in administrative burden on the server and

increasing the availability and scalability of resources that enhance the productivity of IT technicians.^{9–13}

Upon special request, the Centers for Medicare and Medicaid Services (CMS) makes available to researchers' identifiable files,¹⁴ which are data files that require secure storage. This case report from Emory University describes an innovative process of developing and implementing an AWS platform to store and analyze CMS data files.

Background

A research team at Emory University was funded by the Agency for Healthcare Research and Quality (AHRQ) to

Received: January 9, 2023; Revised: September 5, 2023; Editorial Decision: January 4, 2024; Accepted: January 13, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of the American Medical Informatics Association.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

For commercial re-use, please contact journals.permissions@oup.com

examine the care outcomes of Medicare beneficiaries in four states—California, Florida, New Jersey, and Pennsylvania. Data from 2016 included Medicare Provider Analysis and Review (MEDPAR), Outpatient, Carrier (Physician Part B), and Master Beneficiary Summary files. During the data request process, the research team was asked by CMS to demonstrate the feasibility of cloud computing using Medicare claims data. The Emory University data request contained a smaller combined dataset of roughly 3 million hospital discharges, which presented an ideal situation to test CMS research data in a cloud-based environment.

A series of meetings followed with CMS and the Privacy Manager of MBL Technologies, a company providing digital solutions to clients within federal, individual, and commercial entities. The Emory University team included support provided by the Deputy Chief Information Officer from the Office of Information Technology; the Associate Director of Solutions Architecture, a cloud solution engineer from Library & Information Technology Services; the Assistant Dean for Technology (RLS) and the Senior Director of IT Operations from the School of Nursing; and the principal investigator (JPC) of the AHRQ-funded study. These meetings focused on Emory meeting all security and privacy standards outlined by CMS.

CMS security and privacy requirements

Emory University entered a Data Use Agreement with CMS under the CMS Data Privacy Safeguard Program¹⁵ and completed a Data Management Plan Self-Attestation Questionnaire (DMP SAQ)¹⁶ to demonstrate that Emory was prepared to meet all CMS security and privacy requirements. CMS required that the Emory University data custodian complete, sign, and attest through official documentation that the Emory system had in place 18 DMP SAQ required security controls and 8 DMP SAQ required privacy controls that were verified through completion of 113 evidence-based response items.

Materials and methods

AWS at Emory platform

The AWS at Emory platform fulfilled CMS requirements with tailored set of controls built specifically for this project. Emory University's Office of Information Technology (OIT) designed and developed AWS at Emory¹⁷ to provide a HIPAA compliant, secure environment for research-intensive cloud computing. Launched in 2019, the platform integrates with the Emory University management and financial systems. Each Virtual Private Cloud (VPC)¹⁸ available in either the AWS US-East-1 Virginia region or the US-East-2 Ohio region is protected by the Emory firewall, which prohibits access from the public, and connected to Emory's on-premises network with AWS Direct Connect.¹⁹ AWS Service Control Policy,²⁰ AWS Identity Access Management²¹ and Emory Security Risk Detection (SRD) and Security Remediation Service (SRS) implement security controls for all accounts. The SRD and SRS are a series of background processes that continuously check for potential risks in each account and remediate or alert as needed.

After analyzing the CMS Data Management Plan Self-Attestation Questionnaire (DMP SAQ),¹⁶ OIT recognized many of the requirements could be fulfilled with AWS at Emory; however, a few additional technical controls or

manual processes would be required. These included restricting user access to specific resources and data storage, manually reviewing and updating custom software for updates, and implementing additional audit logs and remediation plans. Emory University implemented the needed controls and processes while outlining their function in a system security plan submitted to CMS.

Office of information technology

The Emory University background check process meets the CMS requirements for criminal background checks including fingerprinting,²² and it is used to approve user access to the AWS at Emory account. This process includes Emory rules of behavior to be agreed upon initially and validated annually for all authorized users. The Emory Single Sign-On with MyNetID Group authenticates users to the account. The Emory Active-Directory Lightweight Directory Service (AD LDS)²³ authenticates users accessing the AWS Elastic Compute Cloud (EC2).²⁴ Authorized users have limited access to specifically approved AWS resources through AWS Identity and Access Management²¹ user policies and other AWS service-resource-based access policies. The account's security policies and security controls also block unauthorized users. When a user is no longer associated with the project, account or university, access to the account and its resources are removed as per the Emory University access termination policy.

Initial data management included storage of raw encrypted CMS data within AWS Simple Storage Service (S3)²⁵ with access limited to approved administrators and Emory OIT members. A working version of encrypted data utilized for analyses are stored in a separate S3 bucket that is encrypted at-rest. These data are limited to authorized users who are blocked from accessing any other S3 buckets. All S3 buckets have the versioning feature enabled for better collaboration and tracking. AWS Elastic Block Store (EBS)²⁴ Snapshot backups are enabled with AWS Data Lifecycle Manager (DLM)²⁴ for backup compliance and protection of block storage data.

To generate approved EC2 instances, compute environments are configured by Emory OIT through Amazon Machine Images (AMIs).²⁴ Emory OIT and IT operations maintain patches and update the AMIs on a regular basis to meet all Emory and CMS security and compliance policies. Patches and updates are routinely monitored and pushed out to the account, with monthly reports that are sent to Emory OIT administrators. These AMIs are specifically configured with a custom logon banner to display validity of the environment. Only authorized users can employ AMIs to launch instances after obtaining approval. Ansible Automation Platform²⁶ configures software and tools within EC2 instances with policies blocking non-approved software installations. For analyses, all data are encrypted in-transit and at-rest within the environment.²⁷ For added privacy and security, all EC2 instances have logout and inactivity policies that automatically disconnect the user after 10 failed attempts and block access for 30 minutes, while automatically logging out any inactive user.

The AWS environment and its resources are monitored through comprehensive logging. AWS VPC¹⁸ Flow Logs capture network traffic data by traversing the network interfaces of the VPC and EC2 instances. All S3 buckets have server access logging enabled, and S3 Access Logs are stored in a

separate S3 bucket to track data access. Internal systems clocks of instances are synchronized with a common authoritative time source to align timestamps of all events for consistency and accuracy of EC2 logs. CloudWatch²⁸ is used to securely store EC2 logs and metadata for audits along with User Access and Usage Logs. All logs are retained for 12 months per Emory University policy. All users of the account are blocked from accessing all logs stored in AWS CloudWatch and S3 buckets, only exception being approved administrators and Emory OIT members for tracking and auditing. The AWS CloudTrail²⁹ tracks account application programming interface (API) calls for 12 months, and AWS Athena allows logs to be queried and reviewed. All logs are protected from unauthorized access, deletion, and modification. Authorized Emory University School of Nursing IT personnel follow specific instructions provided to monitor and audit logs.

A security plan was implemented to ensure all security policies are followed. This plan includes documenting configurations of individual hosts within the system, maintenance of security functions, tracking known vulnerabilities through a Plan of Action and Milestones (POA&M),³⁰ and other documentation as needed for the system’s secure operation. The NIST SP 800-160 Volume 1³¹ was referenced to meet CMS requirements for a trustworthy secure system. Information systems use FIPS 140-2 validated cryptographic modules for

transmission of data-in-motion and data-at-rest. EC2 instances are protected from malicious code with up-to-date virus definitions with important file systems scanned every 12 hours and a full system scan every 72 hours. AWS at Emory infrastructure is secured with regular vulnerability scans and reviewing audit logs. If a vulnerability or data breach is identified, from scan findings or approved security assessment or audit report, then all findings are documented in a POA&M reported to the Department of Health and Human Services (HHS) and remediated within the CMS required timelines of 15 days for critical vulnerabilities and 30 days for high vulnerabilities based on the date identified and not the creation date of the POA&M.

Technical implementation

High-level architecture (Figure 1) illustrates the specifications and implementation of the OIT steps described above. Authorized users, properly vetted, can login to the account and utilize the EC2 service to run SAS and STATA from Emory OIT configured AMIs. Once the users are authenticated, they only have access to a specific S3 bucket that contains a working copy of CMS data for analysis. A high-level analysis flow (Figure 2) illustrates the process to analyze the working copy of CMS data. Access and usage logs of the account and CMS data are stored automatically to audit user

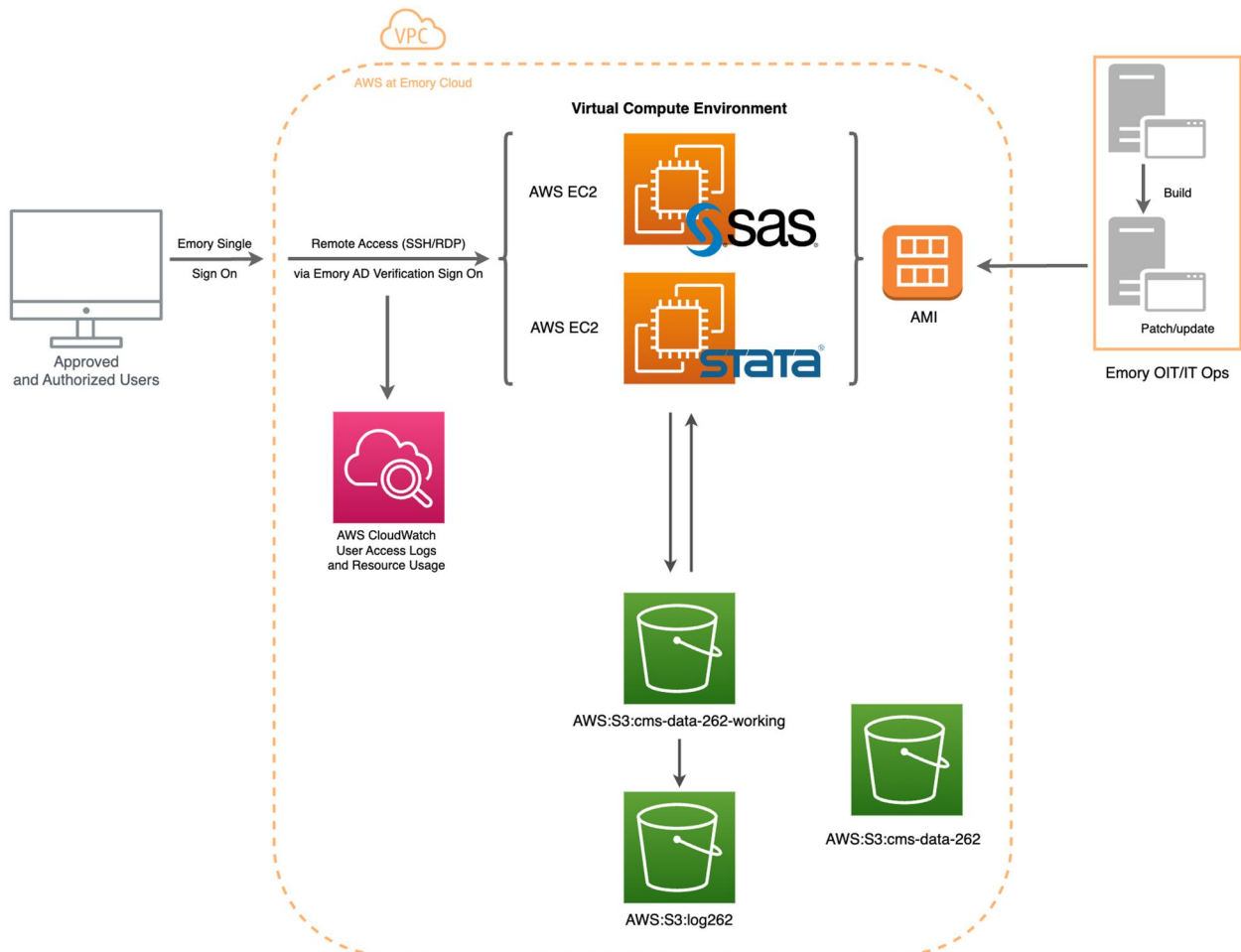


Figure 1. A high-level architecture diagram of the system implemented at Emory University and approved by the Centers for Medicare and Medicaid Services.

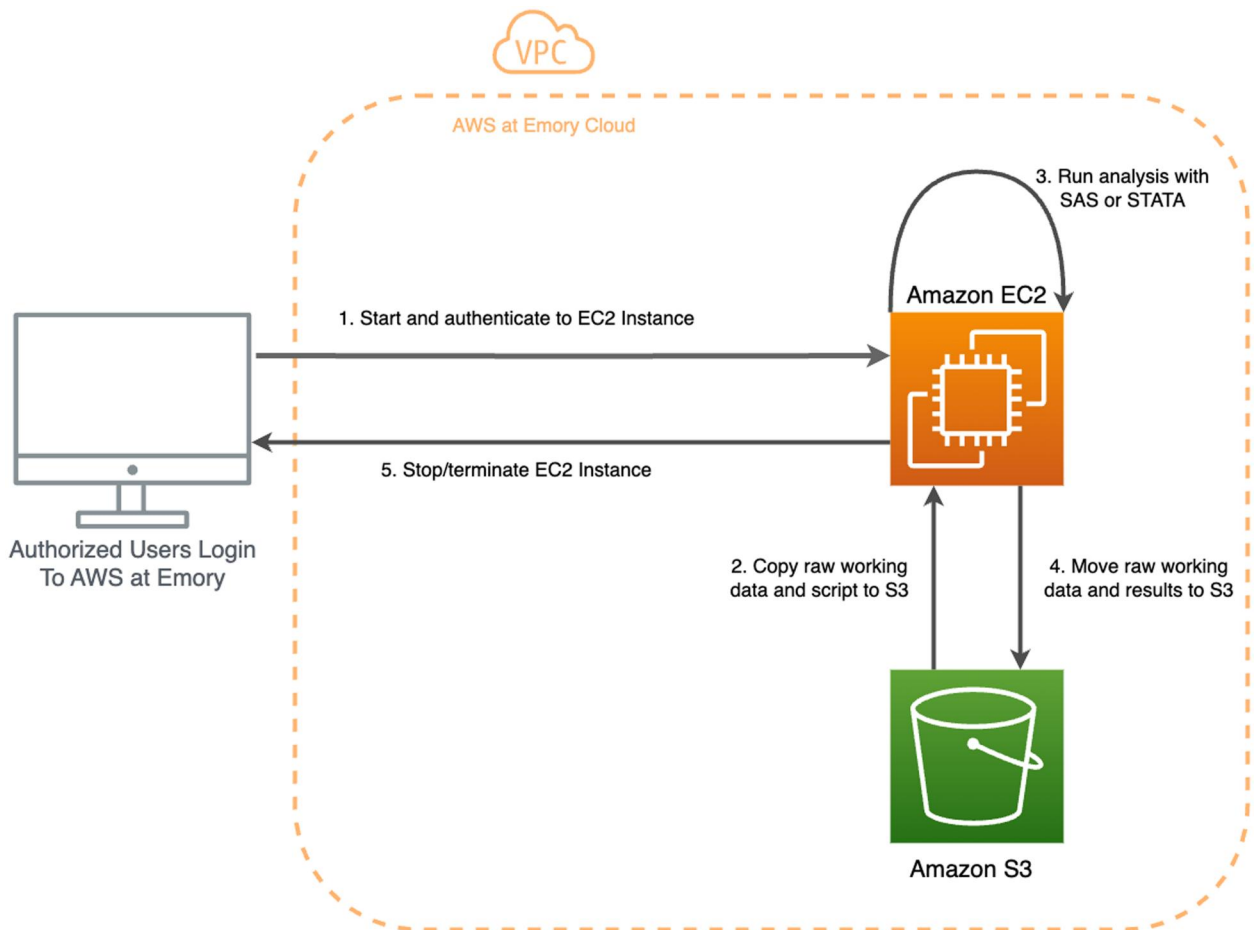


Figure 2. A diagram of the high-level analytic flow architecture available through AWS at Emory and approved by the Centers for Medicare and Medicaid Services.

access and resource usage to ensure secure data access and analysis.

Human information processing was implemented to prevent usability issues and to effectively adopt and utilize cloud services.³² This was achieved through training on use of AWS at Emory platform, user restricted data access, limited or blocked user cloud services based on project needs, and only approved tools to analyze and store results. All authorized users agreed to the rules of behavior to not install any software in their home directories, and School of Nursing (SON) IT performs periodical scans to ensure no unauthorized software is installed. Licensing information of approved software is also maintained by SON IT.

SON IT maintains an up-to-date inventory of laptop computers connecting to EC2s with security patches that are pushed to all connected computers. Monthly reports are used to identify outstanding security patches. These reports are reviewed by SON IT staff who then contact users to update their computer with the latest patches. A follow-up with users and reviewing logs help determine if the patches were applied to each computer. Access termination process was documented by SON IT and the principal investigator of the project to maintain and update user access.

SON IT is responsible for the ongoing review of audit logs documenting any suspicious activities, unauthorized access, abnormal behaviors, and auditing system failures. This includes a weekly review of logs and maintaining documentation of each review.

Conclusion, lessons learned, and future directions

Efficient coordination of IT systems and human processes is required to maintain a secure cloud computing environment. In this case report, the team recognized the essential role of human factors in the maintenance of technology, including but not limited to ongoing support, such as AMIs, software updates, and training. The team observed a complex learning curve associated with training, device security, becoming familiar with foreign terminology and the cloud system interface, and acknowledging variances in launching and stopping instance by operation system (ie, PC vs Mac).³³ In addition, the team experienced unanticipated increases in transition costs when users were not familiar with instance state and run times. Better communication and collaboration among the researchers were observed as researchers' increased cloud computing knowledge. Future work should focus on structured training for researchers without a cloud computing background as essential to facilitate the successful use of cloud computing resources. This experience with CMS guidelines resulted in successful implementation of identifiable data for use by the research team.

Acknowledgements

The authors would like to acknowledge the management and staff from the Office of Information Technology at Emory University for their tireless efforts to ensure the successful implementation of this project.

Author contributions

R.L.S. and J.P.C. contributed to the conceptualization and supervision of this case report. J.A.L. and C.T. designed and provided visualization of the platform. R.L.S. and J.P.C. were the primary authors and Y.J.K. and Y.L. provided revisions. All authors have reviewed this case report, and they have provided final approval of the version to be published.

Funding

This work was supported by the Agency for Healthcare Research and Quality (AHRQ) grant number R01HS026232.

Conflicts of interest

None declared.

Data availability

No new data were generated or analyzed in support of this research.

References

- De Donno M, Tange K, Dragoni N. Foundations and evolution of modern computing paradigms: Cloud, IoT, edge, and fog. *IEEE Access*. 2019;7:150936-150948. doi:10.1109/ACCESS.2019.2947652
- Navale V, Bourne PE. Cloud computing applications for biomedical science: a perspective. *PLoS Comput Biol*. 2018;14(6):e1006144. doi:10.1371/journal.pcbi.1006144
- Alkasem A, Liu H, Zuo D, Algarash B. Cloud computing: a model construct of real-time monitoring for big dataset analytics using apache spark. *J Phys: Conf Ser*. 2018;933:012018. doi:10.1088/1742-6596/933/1/012018
- Samea F, Azam F, Rashid M, Anwar MW, Haider Butt W, Muzafar AW. A model-driven framework for data-driven applications in serverless cloud computing. *PLoS One*. 2020;15(8):e0237317. doi:10.1371/journal.pone.0237317
- Sareen P. Cloud computing: Types, architecture, applications, concerns, virtualization and role of IT governance in cloud. *IJARCSSE*. 2013;3(3):533-538.
- Alvarez RV, Mariño-Ramírez L, Landsman D. Transcriptome annotation in the cloud: complexity, best practices, and cost. *Giga-science*. 2021;10(2):1-11. doi:10.1093/gigascience/giaa163
- Schadt EE, Linderman MD, Sorenson J, Lee L, Nolan GP. Computational solutions to large-scale data management and analysis. *Nat Rev Genet*. 2010;11(9):647-657. doi:10.1038/nrg2857
- Rosenthal A, Mork P, Li MH, Stanford J, Koester D, Reynolds P. Cloud computing: a new business paradigm for biomedical information sharing. *J Biomed Inform*. 2010;43(2):342-353. doi:10.1016/j.jbi.2009.08.014
- Richter F. Amazon, Microsoft & google dominate cloud market. *Statista*. 2022; Accessed January 1, 2023. <https://www.statista.com/chart/18819/worldwide-market-share-of-leading-cloud-infrastructure-service-providers/>
- Sobeslav V, Maresova P, Krejcar O, Franca TCC, Kuca K. Use of cloud computing in biomedicine. *J Biomol Struct Dyn*. 2016;34(12):2688-2697. doi:10.1080/07391102.2015.1127182
- Patel AR, Tiwari RV, Khureshi RA. Comparative study of top cloud providers on basis of service availability and cost. *IJFMR—International Journal for Multidisciplinary Research*. 2022;4(6):1-8. doi:10.36948/ijfmr.2022.v04i06.1140
- Fulmer J. *Top 10 Cloud Providers & Companies* 2022. IT Business Edge; 2021. Accessed December 13, 2022. <https://www.itbusinessedge.com/cloud/compare-top-cloud-providers/>
- Mukherjee S. Benefits of AWS in modern cloud. *SSRN J*. 2019;1-21. doi:10.2139/ssrn.3415956
- Research Data Assistance Center (ResDAC). *Research Identifiable File (RIF) Requests*; 2022. Accessed August 18, 2022. <https://resdac.org/research-identifiable-files-rif-requests>
- Research Data Assistance Center (ResDAC). CMS Data Privacy Safeguard Program (DPSP). Accessed August 18, 2022. <https://resdac.org/articles/cmss-data-privacy-safeguard-program-dpsp>
- Centers for Medicare and Medicaid Services. Data Privacy Safeguard Program Data Management Plan Self-attestation Questionnaire (DMP SAQ). Accessed August 18, 2022. <https://resdac.org/request-form/dmp-saq>
- Wheat RA. Implementing Centrally Managed Cloud Security and Compliance Controls for Amazon Web Services within a Large Research Enterprise and Healthcare System. Emory University. Accessed December 13, 2022. https://it.emory.edu/_includes/documents/sections/aws/EmoryAWS.pdf
- Amazon Web Services. *Amazon Virtual Private Cloud—User Guide*. 2022. Accessed August 18, 2022. <https://docs.aws.amazon.com/pdfs/vpc/latest/userguide/vpc-ug.pdf#what-is-amazon-vpc>
- Amazon Web Services. *AWS Direct Connect—User Guide*. 2022. Accessed July 17, 2022. <https://docs.aws.amazon.com/pdfs/direct-connect/latest/UserGuide/dc-ug.pdf>
- Amazon Web Services. *AWS Organizations—User Guide*. 2022. Accessed July 17, 2022. https://docs.aws.amazon.com/pdfs/organizations/latest/userguide/organizations-userguide.pdf#orgs_introduction
- Amazon Web Services. *AWS Identity and Access Management—User Guide*. 2022. Accessed July 17, 2022. <https://docs.aws.amazon.com/pdfs/IAM/latest/UserGuide/iam-ug.pdf#introduction>
- Centers for Medicare and Medicaid Services (HHS). *Criminal Background Checks*. Vol § 455.434.; 2021, 525.
- Microsoft. *Active Directory Lightweight Directory Services Schema*. August 18, 2022. <https://learn.microsoft.com/en-us/windows/win32/adschema/active-directory-schema>
- Amazon Web Services. *Amazon Elastic Compute Cloud—User Guide for Linux Instances*. 2022. Accessed July 17, 2022. <https://docs.aws.amazon.com/pdfs/AWSEC2/latest/UserGuide/ec2-ug.pdf#Instances>
- Amazon Web Services. *Amazon Simple Storage Service—User Guide*. 2022. Accessed July 17, 2022. <https://docs.aws.amazon.com/pdfs/AmazonS3/latest/userguide/s3-userguide.pdf>
- Red Hat Ansible Automation Platform. Accessed December 28, 2022. <https://www.redhat.com/en/resources/ansible-automation-platform-datasheet>
- Amazon Web Services. *Logical Separation on AWS—AWS Whitepaper*. 2020. Accessed August 18, 2022. <https://docs.aws.amazon.com/pdfs/whitepapers/latest/logical-separation/logical-separation.pdf#introduction>
- Amazon Web Services. *Amazon CloudWatch Logs—User Guide*. 2022. Accessed July 17, 2022. <https://docs.aws.amazon.com/AmazonCloudWatch/latest/logs/cwl-ug.pdf>
- Amazon Web Services. *AWS CloudTrail User Guide*. 2022. Accessed August 18, 2022. <http://awsdocs.s3.amazonaws.com/awscloudtrail/latest/awscloudtrail-ug.pdf>
- Centers for Medicare and Medicaid Services. *Plan of Action and Milestones Process Guide*. 2021. Accessed August 18, 2022. <https://www.cms.gov/files/document/cms-poam-process-guide-v11.pdf>
- Ross R, McEvilly M, Carrier Oren J. *Systems Security Engineering: Considerations for a Multidisciplinary Approach in the Engineering of Trustworthy Secure Systems*, NIST SP 800-160, 2016. National Institute of Standards and Technology. doi:10.6028/NIST.SP.800-160
- Safianu O, Twum F, B J. Information system security threats and vulnerabilities: evaluating the human factor in data protection. *IJCA*. 2016;143(5):8-14. doi:10.5120/ijca2016910160
- Gera A, Xia CH. Learning curves and stochastic models for pricing and provisioning cloud computing services. *Service Science*. 2011;3(1):99-109. doi:10.1287/serv.3.1.99