CLOUD COMPUTING COURSEWORK: TECHNICAL REPORT

“Designing a native cloud application for a medical imaging application”

By student: 21025647

1. Introduction

1.1 Background

Cloud computing facilitates the storage, analysis, and real-time distribution of imaging data, like MRIs, across healthcare facilities. This leads to an increase in accessibility, resulting in accelerated diagnoses and treatment plans, and more effective healthcare. In this report, I create a detailed design and discuss AWS cloud-based solutions for medical imaging. I will investigate and apply cloud service models like IaaS (Infrastructure as a Service), PaaS (Platform as a Service), and SaaS (Software as a Service) as well as cloud-related characteristics like storage, scalability, and security showcasing their integration and utility in building the application.

2. Cloud Solutions, Deployment Models, and Services

2.1 Cloud Computing Solutions

Cloud computing offers many services and resources via the internet through advanced cloud technologies. These solutions have many beneficial advantages (Figure 1) like scalable and flexible access to various computing resources, including storage, processing power, and software applications, thereby reducing the need for extensive physical infrastructure. Notable advancements in this field include the introduction of Amazon Web Services (AWS) by Amazon followed by Google's development of the Google App Engine then Microsoft's Azure (AWS Exam Guide, n.d.)

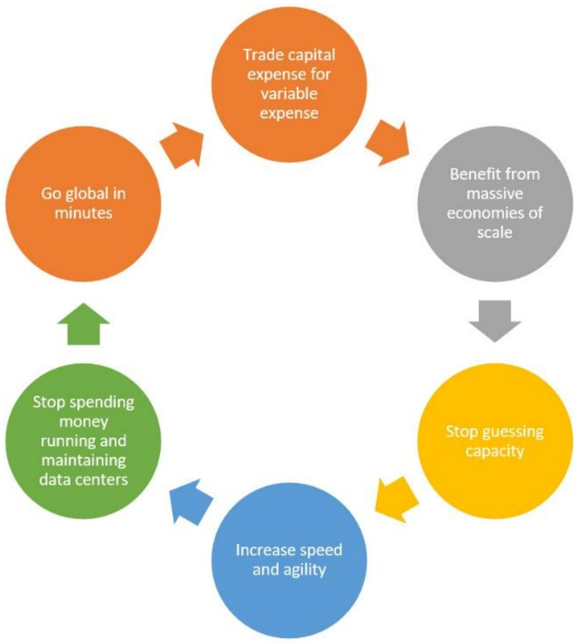

Figure 1: Advantages of cloud computing

2.2 Cloud Models and Services

The deployment of cloud services is categorized into models. Private Clouds are owned and maintained by individual organizations, offering security and control, though they may lack the scalability of their counterparts. Public Clouds, managed by third-party providers such as AWS, Azure, and Google Cloud, strike a balance between cost-effectiveness and scalability but might not offer the same level of customization or security. Hybrid Clouds combine the advantages of both private and public clouds, catering to a range of needs by utilizing private clouds for sensitive data and public clouds for less critical workloads. Community Clouds are specifically tailored for distinct communities or groups with shared interests, providing a unique collaborative platform that allows for efficient resource sharing and data collaboration (Microsoft, n.d.).

Cloud services, classified into Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS), each play a distinct role in the cloud computing ecosystem:

1. IaaS forms the foundation within cloud computing, it provides pay as you go computing resources over the internet. IaaS providers offer virtual servers, service storage, networking infrastructure and other computational resources. Due to these resources being on-demand, users avoid the expense of buying hardware. Amazon Web Services (AWS) and Rackspace are examples of IaaS providers. They both allow customers to rent processing power, service storage, and compute resources (in AWS there's the elastic compute cloud (EC2), simple storage service (S3) and the virtual private cloud (VPC) whereas in Rackspace there's the cloud servers, cloud files and cloud networks). Overall, increasing scalability and flexibility as customers run their cloud-based data centres. In relation to the design, IaaS allows healthcare facilities to scale according to the demands of medical imaging departments in a very secure, cost effective and time efficient way (e.g the more MRI or CT scans required, the more the storage and processing power will increase).

2. PaaS includes all the features of IaaS but in addition offers a development suite. Allowing for the development, implementation, and management of applications, ultimately simplifying aspects of software development. Microsoft Azure and Google App Engine are examples of PaaS providers. They both allow customers to build, test, deploy, and manage applications efficiently, these platforms provide IaaS in addition to middleware, development frameworks, and tools for coding, debugging, and testing applications. In relation to the design, PaaS provides tools that can be used to manage patient records, schedule appointments, and allows for the integration of diagnostic tools in a regulatory compliant way.

3. SaaS includes off the features of the other 2 models but additionally offers the hosting and maintenance of the software, handling of the infrastructure, and ensures the availability and security of the app. Meaning users need not worry about complex software and hardware management. Salesforce for customer relationship management (CRM) and Google Workspace for productivity tools are examples of

SaaS providers. Salesforce offers a suite of CRM services, allowing businesses to manage and analyse customer interactions and data throughout the customer lifecycle. Similarly, Google Workspace provides different productivity applications such as Gmail, Docs, and Drive, facilitating seamless collaboration and communication for users. In relation to the design, SaaS helps use software for patient response and interactive management and electronic health records (EHRs) without the need for in-house development or extensive IT infrastructure (AWS Exam Guide, n.d.).

3. Cloud Design and Implementations

3.1 Suggested Overview Design

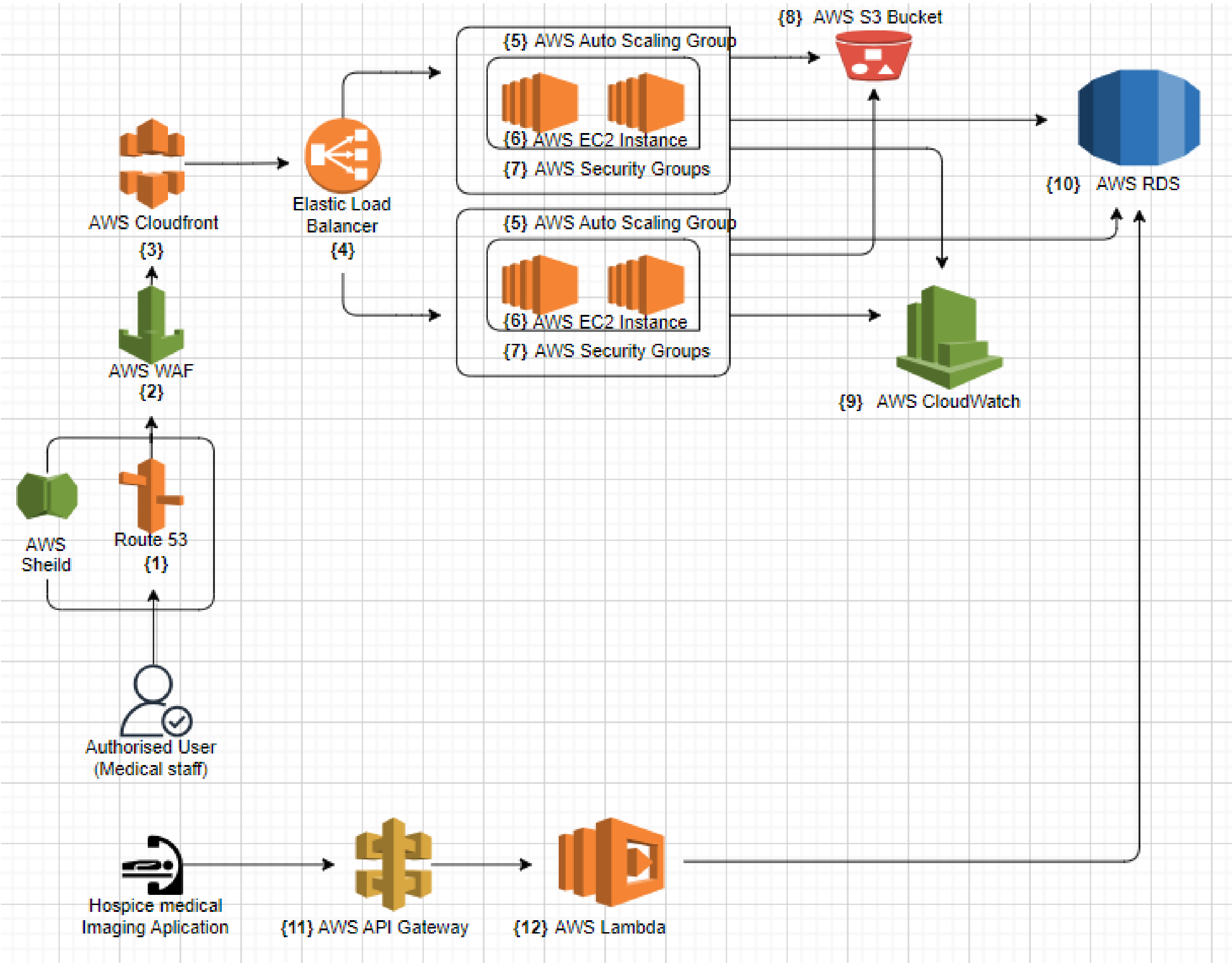

Figure 2: Proposed AWS Cloud Architecture for Hospice Medical Imaging Application. Including network setup with Route 53 for DNS services, AWS CloudFront for content delivery, and AWS WAF for security. It is depicted using AWS Auto Scaling Groups and EC2 Instances for computational capacity, RDS for relational data storage, S3 Bucket for object storage, CloudWatch for monitoring, and the integration of AWS API Gateway and Lambda to create a Restful API management. Built using Draw.io.

3.2 Design Description

This diagram uses AWS services to provide a robust, scalable, and secure solution for medical imaging department in a hospital. Starting with AWS Route 53 for Domain Name System (DNS) management and traffic routing, it leverages AWS Shield for Distributed Denial of Service (DDoS) protection and AWS Web application Firewall (WAF) to safeguard against web exploits. AWS CloudFront serves as a Content Delivery Network (CDN) to reduce latency, while the Elastic Load Balancer ensures even traffic distribution across the EC2 instances within an Auto Scaling Group to handle load efficiently. Security Groups protect these instances, which host the application server and APIs. AWS S3 bucket provides durable storage for medical images, and AWS Relational Database Service (RDS) is used for data storage. AWS CloudWatch is used for monitoring the application, AWS API Gateway acts as the front door for the serverless functions hosted on AWS Lambda, offering a serverless approach for application components and backend services and acting as a RESTful API. The integration AWS products significantly enhance the efficacy of this application.

3.3 Integration of AWS products

3.3.1 Web Design*:*

{1} AWS Route 53 directs authorised staff and systems to the application, allowing them to find and access it. Connected to the AWS Shield which gives extra protection against DDos attacks.

{2} AWS WAF keeps images and data safe from unauthorized users and cyber-attacks.

{3} AWS CloudFront speeds up the delivery of the images to staff.

{4} Elastic Load Balancer makes sure the application doesn't get overwhelmed by spreading out requests for images and data.

{5} AWS Auto Scaling Group helps the application work smoothly and keep up with demand, adjusting the number of servers being used.

{6} AWS EC2 Instances are the computers where the medical imaging application runs and processes images.

{7} AWS Security Groups are the security guards of the application, setting the rules for who can send data to and from the EC2 computers, keeping patient information more secure.

{8} AWS S3 Bucket is the secure place where all the medical images and related data are stored.

{9} AWS CloudWatch spot and report any issues with the application.

{10} AWS RDS is the database that holds patient data and other structured data, which can be accessed at any time.

{11} AWS API Gateway is the main entrance of the imaging application that doctors can use to send requests for patient images and data is a secure way.

{12} AWS Lambda does tasks (like analysing) quickly and on demand without needing a permanent server (Amazon, 2023).

3.3.2 REST Interface for Access and Search options

When staff need to access or search for patient data or medical images, they send an HTTP request to **{11}** the AWS API Gateway. This request will have parameters, like a patients NHS number. After getting the request, **{11}** routes it to the **{12}** AWS Lambda function. Lambda works with other AWS services to fulfil the request. It retrieves structured patient data from **{10}** the AWS RDS (10). Then Lambda queries image metadata form **{8}** the AWS S3 bucket. The data collected is then usually formatted in a JSON structure to make interpretation for the application easier. The final response gets sent back through **{11}** the AWS API Gateway to the healthcare professional's application, providing the requested data for medical investigation. This can be used to access image metadata, patient results and patient information (AWS Exam Guide, n.d.).

3.3.3 Storage, Scalability and Security

Storage uses **{8}** Amazon S3 (highly scalable and durable), as the object storage component. Each patient can have a dedicated S3 bucket, with strict controls to ensure data protection and privacy. Medical images are uploaded to the S3 buckets and stored until staff request for access. The use of **{10}** the AWS RDS is also used for structured data management, replacing what some would use the DynamoDB for. The RDS is highly reliable and scalable and allows for the effective storing and management of patient records and other relational data. traditional SQL queries, done through **{11}** the AWS API Gateway are used to retrieve this data.

Scalability can be found using **{5}** AWS Auto Scaling, which automatically adjusts the number of **{6}** EC2 instances to coincide workload changes without manual intervention. **{4}** The Elastic Load Balancer spreads incoming requests across the instances, making sure that any one single server has too much to handle. With data being backed up regularly from **{8}** AWS S3 and **{10}** AWS RDS, if a situation required quick restoration after data loss. This setup ensures that the application is accessible and able to scale. Even in the event of unforeseen situations.

Security has been taken very seriously as this medical application will harbour confidential data. **{7}** AWS Security Groups provide firewall rules (which can be customised) for the **{6}** EC2 instances, allowing control of inbound and outbound traffic. Data in transit is protected by Secure Sockets Layer (SSL) or Transport Layer Security (TLS), securing HTTP traffic between staff and the imaging application. There are also AWS's built-in encryption options to secure data at rest within the **{8}** S3 buckets where medical images are stored and the **{10}** the RDS instances that manage patient data (Vyas, 2023).

3.3.4 Other Aspects for Consideration

The AWS's pay-as-you-go model means it is extremely cost effective as healthcare facilities only pay for the resources they consume. For instance, AWS services like **{6}** EC2, **{12}** Lambda, **{10}** RDS, and **{8}** S3 offer scalable costs in line with usage.

Maintenance is another crucial aspect. **{9}** AWS CloudWatch plays a role in monitoring the application's performance and logging system problems. This tool is a proactive issue detector and debugger. Despite this, regular updates will still be necessary for security and functionality reasons (Obregon, 2023).

3.3.5 Conclusion

In Conclusion, this report effectively critically analyses the practical application of IaaS, PaaS, and SaaS concepts using AWS resources. The integration of AWS in the medical imaging shows the efficiency of cloud computing models in healthcare technology. This design shows many of cloud computing’s benefits including scalability, security, a way for managing sensitive medical data and a REST interface for accessing various types of data including patient details, diagnosis results, and image metadata. The proposed application aligns with the best practices in cloud architecture and healthcare IT standards.

4. Bioinformatics Application

4.1 Hadoop and MapReduce

Several challenges were encountered in the development of this application, leading to its incomplete status. If the project had progressed as planned, the following steps would have been undertaken:

The initial phase involved uploading seven .gaf files to an AWS S3 bucket. The goal was to identify organisms exhibiting the 'competence' feature, particularly focusing on the GO:0030420 term and its related sub terms. The expected outputs from this are expected to include comprehensive reports with statistical analysis and visualizations, along with lists of proteins and their GO terms for each organism. As well as comparative data showing common and unique GO terms among the organisms. These outputs would have been stored in the S3 bucket for efficient retrieval and furthers use.

In terms of design, the project centred around the use of Hadoop and MapReduce for efficient and distributed data processing. This approach is generally effective for managing large datasets. A Java program named 'GoTerm-Count.java' could've been developed for processing the files to extract relevant GO terms. The integration of AWS services like the S3 bucket for storage, AWS Lambda for processing, and AWS Athena for data warehousing was considered due to their scalability and cost-effectiveness. However, the design of the MapReduce program was to assume tab-separated input files with the GO ID in a specific field, which might limit its ability to have different data formats.

4.2 Conclusion

To address these challenges, resolving compatibility and permission issues with Windows and Cygwin was identified as a necessity. Future repeats of this project would benefit from accommodating various data formats to make sure data integrity is maintained. Despite the technical constraints that prevented the completion of the application, the design and planned implementation demonstrated the potential for a scalable and efficient solution for analysing the bacterial organisms.

5. References

Amazon (2023). Amazon Web Services (AWS) - Cloud Computing Services. [online] Amazon Web Services, Inc. Available at: https://aws.amazon.com/.

AWS Exam Guide (n.d.). AWS Certified Cloud Practitioner Exam Guide. [online] subscription.packtpub.com. Available at: https://subscription.packtpub.com/book/cloud-and-networking/9781801075930/2/ch02lvl1sec07/summary.

Draw.io (n.d.). Flowchart Maker & Online Diagram Software. [online] app.diagrams.net. Available at: https://app.diagrams.net/.

Microsoft (n.d.). Public Cloud vs Private Cloud vs Hybrid Cloud | Microsoft Azure. [online] azure.microsoft.com. Available at: https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-are-private-public-hybrid-clouds.

Obregon, A. (2023). Getting Started with AWS: A Beginner's Guide. [online] Medium. Available at: https://medium.com/@AlexanderObregon/getting-started-with-aws-a-beginners-guide-d4f7542a9840.

Vyas, V. (2023). An In-depth Guide to Amazon S3: Secure, Scalable, and Simple Storage. [online] Medium. Available at: https://vipulvyas.medium.com/an-in-depth-guide-to-amazon-s3-secure-scalable-and-simple-storage-70a6853313a5?source=rss-------1.

www.youtube.com. (n.d.). AWS end to end Architecture for Web App, web services and database. [online] Available at: https://www.youtube.com/watch?v=fCIDssg1oeU .

www.youtube.com. (n.d.). Cloud Computing Services Models - IaaS PaaS SaaS Explained. [online] Available at: https://youtu.be/36zducUX16w?si=jKSua5RZQHtNT4JY.

www.youtube.com. (n.d.). MapReduce Tutorial | MapReduce In Hadoop | MapReduce Example | MapReduce For Beginners | Simplilearn. [online] Available at: https://www.youtube.com/watch?v=fpsa2jtCKmQ .

AWS Academy Cloud Foundations course

Module Number	Lesson Topic	Grade Achieved
Module 1	Cloud concepts overviews	100/100
Module 2	Cloud economics and billings	80/100
Module 3	AWS global Infrastructure overview	90/100
Module 4	AWS cloud security	90/100