ENTERPRISE CLOUD AND INFRASTRUCTURE AUTOMATION 1

MODULE CODE: COMP50008

COMPREHENSIVE CLOUD-BASED SOLUTIONS FOR HOSTING E-COMMERCE WEBSITE

Individual Assignment: 23080330

#Word Count: *1619*

Table of Contents

I. Introduction and Problem Overview

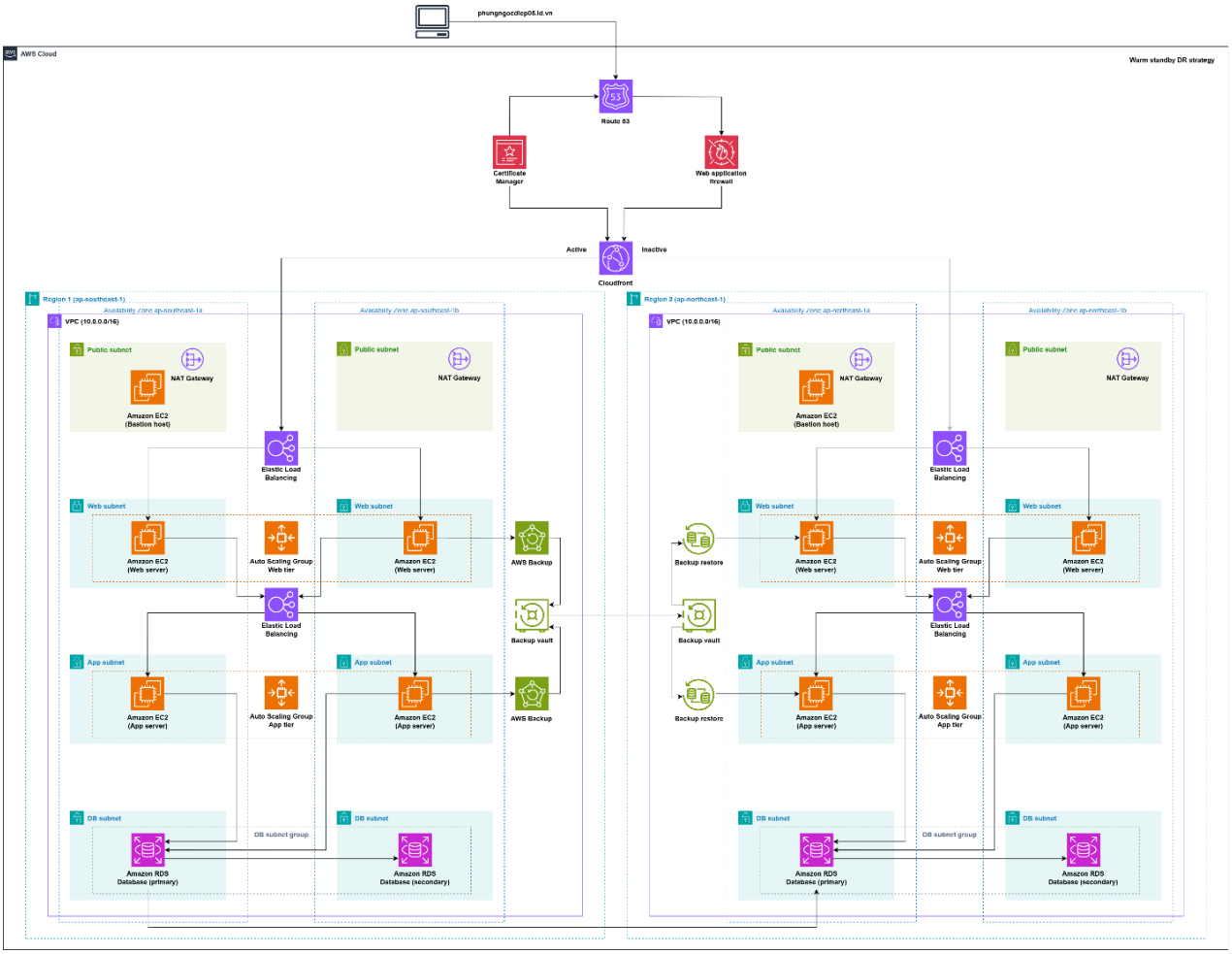
Along with the rapid evolution of technology applications in the digital transformation era across numerous industries, cloud computing also contributes a pivotal role in business management and transformation in innovation operating systems. Recognizing the potential, one owner of Vietnamese clothing stores decides to create an online website and sell his products but faces challenges in selecting suitable cloud-based services for hosting solutions.

The upcoming part of the report will demonstrate a 3-tier architecture e-commerce website hosting with a warm standby Disaster Recovery (DR) strategy in Amazon Web Services (AWS) to ensure zero downtime, data protection, and low latency for a global audience (AWS, 2021). Besides, there will be accompanied basic requirements features like assigning the domain name for the platform, recommended virtual server selection, and scalability to handle high traffic during peak periods, such as festivals. Finally, the report will mention an estimated month-end cost for the architect and the comparison between different database engines in AWS.

II. Proposed Cloud-Based Solutions Architecture

1. AWS architecture diagram

The diagram below illustrates a 3-tier architecture for hosting an e-commerce website on AWS, coupled with a warm standby DR strategy. The architecture features a Presentation Layer (Web Tier), Application Layer (App Tier), and Database Layer (DB Tier) distributed across primary and DR regions. (Jodhani, n.d.)



Here's an architecture diagram description for the components:

Amazon VPC enables the launch of various AWS resources within a custom-defined virtual network (AWS, 2024). The architecture above consists of two VPCs, each configured with eight subnets, two route tables, one NAT gateway, one internet gateway, and four security groups.

Amazon EC2 is a virtual server in the AWS cloud environment (AWS, 2024). The architecture requires two bastion host servers, each configured in the public subnet of their respective regions. These bastion hosts are configured to allow access to the instances of the web private subnet and app private subnet only via SSH protocols, and access is restricted to MyIP only.

Moreover, for security purposes, only the instances in the app subnet are allowed to reach the databases.

Elastic Load Balancing supports Application Load Balancers in distributing incoming traffic across multiple EC2 instances in two different Availabilities Zones (AZs) which prevents any single instance from being overwhelmed and enhances application performance (AWS, 2024).

Auto Scaling Group ensures the adjustment in the number of EC2 instances according to traffic demand. During peak periods such as holidays, if the requests to buy clothes rapidly increase, it will automatically launch additional instances to handle the load. Conversely, it will terminate instances during low-demand periods (AWS, 2024).

Amazon RDS for MySQL is being configured in the database's private subnet for security purposes. For backup and DR purposes, a primary database replicates to two DB instances via read replica. One read replica is deployed in a different Availability Zone (AZ) within the same VPC, while the other is located in a different region to support the disaster recovery strategy (AWS, 2024).

Amazon Route 53 is the service that manages the DNS for the domain and it ensures the user requests are routed to the correct resource. In this architecture, Route 53 has a type A record that directs traffic to the CloudFront Distribution. Moreover, Route 53 can connect to the domain name through NS record with AWS managed nameservers (AWS, 2024).

Amazon CloudFront is a Content Delivery Network (CDN) service that is fully managed by AWS. CloudFront is designed to enhance the performance, availability, and scalability of websites, applications, and APIs by caching content at AWS edge locations globally (AWS, 2024).

AWS Web Application Firewall (WAF) protects a website's resources from various injections of web threats and vulnerabilities by setting rules based on users' demands. The architecture above shows that WAF is attached between Route 53 and CloudFront which means it enhances security by inspecting income traffic and blocking malicious ones if it try to reach the resources (AWS, 2024).

AWS Backup automates and centralizes backup processes across AWS services through the setting of a backup plan. This service setup in the architect ensures that the main resources are backed up daily in order to regenerate a similar fully functional architecture in another region if that in the primary region fails (AWS, 2024).

AWS Certificate Manager manages SSL/TLS certificates for secure HTTPS connections. Because of using CloudFront, ACM has to be requested in the US East (N. Virginia) Region (AWS, 2024).

2. Key Feature of the architecture

2.1 Ensuring Zero Downtime

Zero downtime can be achieved by implementing a resilient architecture that leverages multi-AZ deployments to eliminate single points of failure. In this setup, applications and databases are distributed across multiple fault domains, ensuring that operations can continue seamlessly even if one domain experiences an outage. This architecture forms the backbone of a highly available system, offering redundancy and fault tolerance. (AWS, 2024)

Enhancing this resilience further involves adopting a warm standby disaster recovery (DR) strategy. A warm standby approach ensures the existence of a fully functional replica of the production environment in a secondary region. This configuration minimizes recovery time since the workload remains operational in the secondary region, ready to handle traffic during a failover. To implement this, Amazon Route 53 can be configured to direct traffic to both regions, keeping the primary region active while the secondary region remains passive until activated. Additionally, AWS Backup can automate the replication of AMIs and other critical data from the primary region to the secondary region, ensuring backup files are updated daily and readily available in the event of a disaster. (AWS, 2024)

By integrating multi-AZ deployments with a warm standby DR strategy, organizations can establish a robust infrastructure that ensures business continuity and uninterrupted service. This approach not only minimizes data loss and downtime but also provides a reliable foundation for high availability, even in the face of significant failures.

2.2 Handling Peak Traffic

To effectively manage traffic surges during peak seasons, such as festivals where customer demand for clothing spikes, leveraging a combination of Elastic Load Balancers (ELB), Auto Scaling Groups, and Application Load Balancers (ALB) is essential for seamless performance and scalability.

The Elastic Load Balancer (ELB) dynamically distributes incoming traffic across multiple servers, preventing any single server from becoming overloaded and ensuring high availability during periods of increased demand. This even distribution of traffic enhances reliability and maintains smooth operation regardless of traffic intensity. (AWS, 2024)

With Auto Scaling Groups, the infrastructure automatically adjusts the number of running EC2 instances in real-time based on traffic patterns. A configuration with a minimum of two instances and a maximum of four instances ensures cost efficiency while maintaining sufficient resources to handle fluctuations. This approach guarantees optimal performance and avoids unnecessary expenditures by scaling up or down as needed. (AWS, 2024)

Application Load Balancers (ALB) provide intelligent traffic routing based on HTTP/HTTPS protocols, optimizing resource utilization. ALBs support advanced routing features, such as

directing requests to specific services or instances based on URL paths, headers, or content, further enhancing system efficiency and performance.

Combined, these AWS services deliver a flexible and resilient infrastructure capable of handling unpredictable traffic spikes. By dynamically scaling resources and optimizing traffic distribution, this architecture ensures high availability, superior performance, and cost-effective operation, making it an ideal solution for e-commerce platforms with fluctuating traffic throughout the year.

2.3 Global Content Delivery with Minimal Latency

Delivering content globally with minimal latency is essential for ensuring a seamless user experience, particularly for applications catering to diverse international audiences. AWS CloudFront, a fully managed Content Delivery Network (CDN), plays a pivotal role in achieving this objective. By caching content across more than 400 edge locations worldwide, CloudFront enables users to retrieve data from servers geographically closer to them. This approach significantly reduces latency, accelerates page load times, and ensures consistent application performance, even under varying traffic conditions (AWS, n.d.).

The integration of CloudFront with Amazon Route 53 further optimizes global content delivery. Route 53 enhances routing efficiency by directing user requests to the nearest CloudFront edge location, based on policies such as latency-based or geolocation routing. Combined with built-in security features like AWS WAF (Web Application Firewall), organizations can safeguard applications from common threats while ensuring uninterrupted service.

This architecture offers a secure, scalable, and reliable framework for delivering high-performance, low-latency content worldwide. By leveraging these AWS solutions, businesses can enhance user satisfaction, accommodate growing traffic demands, and maintain operational resilience, making it an ideal choice for modern applications with a global reach.

III. Cost Estimate Calculation

The AWS cost estimate calculation for hosting a cloud-based e-commerce solution is shown in the below table.

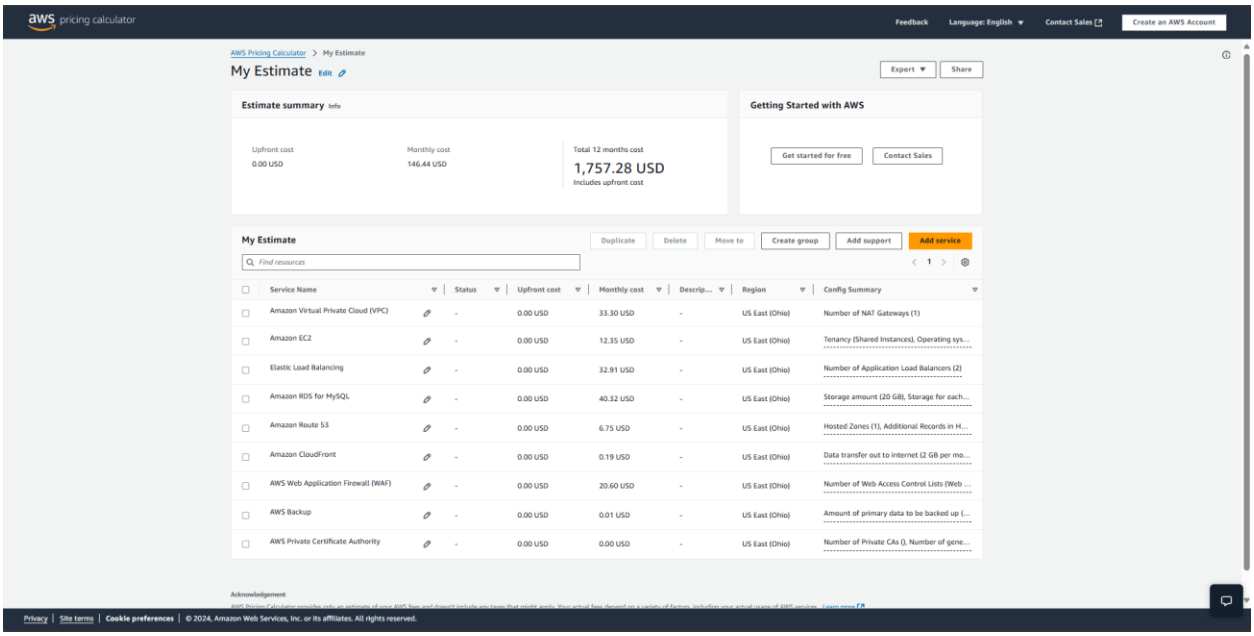


Table 1. Estimate Calculaton (AWS, n.d.)

Based on the client's requirements, the recommended EC2 instances are shared t3.micro instances running the Linux operating system, featuring 2 vCPUs and 1 GiB of memory. An Auto Scaling Group is configured to handle monthly traffic spikes, particularly during week-long promotional periods for festivals. The setup includes two Elastic Load Balancers (ELBs) strategically placed to manage traffic across frontend and backend servers. For database needs, an RDS MySQL instance is suggested, providing a fully managed relational database with a storage capacity of 20 GB.

The projected cost for maintaining this 3-tier e-commerce website is approximately $150 per month, translating to an annual expense of $1,757.28 for a twelve-month period. However, this estimate does not encompass additional data transfer fees or auxiliary services required for maintaining operational excellence.

With a strong emphasis on scalability, security, and low latency, this AWS-based architecture provides a cost-efficient solution tailored to the client's needs. It ensures dynamic scaling during periods of high demand, delivering a seamless user experience for customers worldwide. Furthermore, the secure and robust infrastructure supports the e-commerce platform's reliability and performance, aligning with the client's business objectives.

IV. Comparison of different databases available on AWS

There are various database services that are provided and managed by AWS. Depending on performance, scalabilities, use cases and other characteristics of some databases like Amazon Aurora, PostgreSQL, or MySQL will be suitable in different scenarios.

The table below illustrates the comparison between five different database engines that have been provided by AWS, focusing on their best use cases, performance and scalability.

Database	Type	Best Use Case	Performance	Scalability	Cost
RDS MySQL	Relational	Traditional relational applications	Standard	Vertical, read replica	Moderate
Amazon Aurora	Relational	High-performance workloads	5x MySQL performance	Automatic scaling	Higher than RDS
DynamoDB	NoSQL	Real-time, high-throughput apps	Single-digit millisecond	Unlimited horizontal scaling	Variable (per usage)
Amazon Redshift	OLAP/ Data Warehouse	Analytics and reporting	Optimized for queries	Scales to petabytes	Higher for big data
RDS PostgreSQL	Relational	Applications needing advanced queries	Robust with JSON/ complex queries	Vertical, read replica	Moderate

Table 2: Comparison table among different database engines from AWS (AWS, n.d.)

In summary, the comparison provides overview of numerous database engines, each business will have unique circumstances to selecting the most suitable one based on their needs and operational priorities.

References

AWS, D., 2021. *AWS Serverless Multi-Tier Architectures with Amazon API Gateway and AWS Lambda.* [Online]
Available at: https://docs.aws.amazon.com/whitepapers/latest/serverless-multi-tier-architectures-api-gateway-lambda/introduction.html

AWS, D., 2024. *Amazon CloudFront.* [Online]
Available at: https://docs.aws.amazon.com/AmazonCloudFront/latest/DeveloperGuide/Introduction.html
[Accessed 19 12 2024].

AWS, D., 2024. *Amazon EC2 Auto Scaling.* [Online]
Available at: https://docs.aws.amazon.com/autoscaling/ec2/userguide/auto-scaling-groups.html
[Accessed 19 12 2024].

AWS, D., 2024. *Amazon Elastic Compute Cloud.* [Online]
Available at: https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/Instances.html
[Accessed 18 12 2024].

AWS, D., 2024. *Amazon Relational Database Service.* [Online]
Available at: https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Welcome.html
[Accessed 19 12 2024].

AWS, D., 2024. *Amazon Route 53.* [Online]
Available at: https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/Welcome.html
[Accessed 20 12 2024].

AWS, D., 2024. *Amazon Virtual Private Cloud.* [Online]
Available at: https://docs.aws.amazon.com/vpc/
[Accessed 19 12 2024].

AWS, D., 2024. *AWS Backup.* [Online]
Available at: https://docs.aws.amazon.com/aws-backup/latest/devguide/whatisbackup.html
[Accessed 20 12 2024].

AWS, D., 2024. *AWS Certificate Manager.* [Online]
Available at: https://docs.aws.amazon.com/acm/latest/userguide/acm-overview.html
[Accessed 18 12 2024].

AWS, D., 2024. *AWS WAF, AWS Firewall Manager, and AWS Shield Advanced.* [Online]
Available at: https://docs.aws.amazon.com/waf/latest/developerguide/waf-chapter.html
[Accessed 18 12 2024].

AWS, D., 2024. *Disaster Recovery of Workloads on AWS: Recovery in the Cloud.* [Online]
Available at: https://docs.aws.amazon.com/whitepapers/latest/disaster-recovery-workloads-on-aws/disaster-recovery-options-in-the-cloud.html
[Accessed 20 12 2024].

AWS, D., 2024. *Elastic Load Balancing.* [Online]
Available at: https://docs.aws.amazon.com/elasticloadbalancing/latest/userguide/what-is-load-balancing.html
[Accessed 18 12 2024].

AWS, D., 2024. *Reliability Pillar.* [Online]
Available at: https://docs.aws.amazon.com/wellarchitected/latest/reliability-pillar/shared-responsibility-model-for-resiliency.html
[Accessed 20 12 2024].

AWS, n.d. *Amazon CloudFront Key Features.* [Online]
Available at: https://aws.amazon.com/cloudfront/features/?whats-new-cloudfront.sort-by=item.additionalFields.postDateTime&whats-new-cloudfront.sort-order=desc
[Accessed 18 12 2024].

AWS, n.d. *Amazon Relational Database Service.* [Online]
Available at: https://docs.aws.amazon.com/AmazonRDS/latest/UserGuide/Welcome.html
[Accessed 19 12 2024].

AWS, n.d. *AWS Pricing Calculator.* [Online]
Available at: https://calculator.aws/#/

Jodhani, A., n.d. *Building a Resilient Three-Tier Architecture on AWS.* [Online]
Available at: https://contra.com/p/gFLkRFYg-building-a-resilient-three-tier-architecture-on-aws-with-deploy

Appendix: Original work without AI usage

2.1 Ensuring Zero Downtime

Zero downtime can only be achieved by using a resilient architecture that employs multi-AZ deployments to avoid single points of failure. The applications and databases should be spread over multiple fault domains, each of which can continue operating when the others go down.

Applying a warm standby disaster recovery (DR) strategy further enhances resilience. A warm standby approach guarantees there is exiting a fully function replica of the production environment in another region. This approach ensures the recovery time because the workload always-on in another region. To achieve this approach, there will need a setup in Route 53 where the domain name will connect to both regions but just keep the main region link active while others remain inactive. Besides, to ensure the backup routine, it will need a backup rule to replicate the AMIs of the main instance from the primary region to the secondary region in AWS Backup. Setting the backup plane with the backup rule of working on backup files daily, might ensure the website data is up-to-date when the disaster happens. Moreover, by intergrating multi-

AZ deployments with this DR strategy, organizations can build a robust infrastructure that guarantees uninterrupted service delivery, even in the face of significant failures.

2.2 Handling Peak Traffic

To effectively manage traffic during the peak season such as festivals due to the demand of buying clothes of the customer, leveraging a combination of Elastic Load Balancers (ELB), Auto Scaling Groups, and Application Load Balancers (ALB) ensures seamless performance and scalability.

The ELB will dynamically distribute incoming traffic across multiple servers to ensure no single server is dealing with overload status or maintaining high availability during peak demand. With the configuration of the Auto Scaling Groups, it will automatically adjust the number of running EC2 instances based on real-time traffic patterns with the setup of is maximum of 4 instances and a minimum of 2 instances in the running state. This will ensure the cost optimization while maintaining the optimal resource allocation.

Application Load Balancers further enhance this architecture by routing traffic intelligently based on HTTP/HTTPS protocols. ALBs offer advanced routing features, such as directing requests to specific services or instances based on content, headers, or path patterns, optimizing resource utilization, and improving overall performance.

Altogether, these components from AWS create flexible services that effectively handle unpredictable traffic spikes and ensure high availability while minimizing costs by dynamic scalability according to demand. It is an ideal solution for e-commerce websites that record fluctuating traffic patterns throughout the years.

2.3 Global content with minimal latency

The ability to serve content worldwide with minimal latency is critical in ensuring a seamless user experience, especially for applications that serve diverse, international audiences. AWS CloudFront is a fully managed Content Delivery Network (CDN) and forms a cornerstone solution in this regard. It caches content at more than 400 edge locations around the world, thus allowing users to fetch data from servers closer to their geographical location. This greatly reduces latency, speeds up page load times, and ensures consistent application performance.

By integrating CloudFront with Route 53, together with native security features like AWS WAF, an organization is able to deliver high-performance, low-latency content in a secure and reliable manner across the globe. This architecture also enhances user satisfaction and helps with scalability, reliability, and resilience, making it ideal for modern, globally accessible applications.

III. Cost Estimate Calculation

According to the requirements of clients, the recommended EC2 instances are shared instances that run the Linux Operating System, t3.micro with 2vCPU and 1GiB Memory. The Auto Scaling Group is assigned monthly spike traffic with a duration of a week for stores to run

advertising during festivals along with two ELBs attached to the frontend and backend servers. The recommended database instance is RDS MySQL which fully managed relational database with a capacity of 20GB.

The total monthly cost to run a 3-tier e-commerce website is shown that around 150 USD per month and 1757.28 USD for a twelve-month cost.

However, the estimate still does not account for data transfer costs and other services to maintain operational excellence. With a focus on scalability, security, and minimal latency, this AWS-based architecture balances cost-efficiency with performance, meeting the client's business requirements. These costs ensure the e-commerce platform scales dynamically during high-demand periods and delivers an experience for users worldwide while maintaining a secure infrastructure.