УДК 004.75
DOI https://doi.org/10.36994/2788-5518-2025-01-09-13

DISASTER RECOVERY IN CLOUD COMPUTING SERVICES. AWS GLOBAL ACCELERATOR USAGE FOR A DISASTER RECOVERY SCENARIO

Demydov A. S., *Ph.D. student, Open International University of Human Development "Ukraine", Kyiv, Ukraine. https://orcid.org/0009-0009-7986-9203, anton89d@gmail.com*

Datsenko I. P., *Ph.D., National Defense University of Ukraine, Kyiv, Ukraine. https://orcid.org/0000-0002-0047-413X, docik_ivan@ukr.net*

Abstract. *In the contemporary world of cloud computing, ensuring continuous operation and rapid recovery from disasters is critically important for many organizations. This article examines the use of AWS Global Accelerator as an effective tool for providing disaster recovery in cloud networks. AWS Global Accelerator enhances the availability and performance of applications by routing traffic through AWS's global network, ensuring a more stable and reliable connection.*

The study covers an analysis of the key functionalities of AWS Global Accelerator, including high availability support, latency reduction, and increased data transfer speeds. Special attention is given to disaster recovery scenarios where Global Accelerator enables swift traffic redirection to backup resources in case of a primary network component failure. This significantly reduces downtime and minimizes data loss, which is crucial for maintaining business continuity.

The article also presents experimental research results demonstrating the effectiveness of AWS Global Accelerator in real-world conditions. It is shown that the implementation of this tool can reduce the average recovery time after a disaster by 30%, significantly enhancing the overall resilience of cloud infrastructure. Recommendations for optimal configuration and usage of AWS Global Accelerator are provided to achieve maximum results in various disaster recovery scenarios.

The conclusion highlights the advisability of using AWS Global Accelerator as an integral component of disaster recovery strategies in cloud networks, ensuring high reliability and efficiency of modern information systems.

Key words: *disaster recovery; AWS global accelerator; cloud computing.*

АВАРІЙНЕ ВІДНОВЛЕННЯ В ХМАРНИХ МЕРЕЖАХ. ВИКОРИСТАННЯ AWS GLOBAL ACCELERATOR У СЦЕНАРІЇ АВАРІЙНОГО ВІДНОВЛЕННЯ

Антон Демидов, *аспірант, Відкритий міжнародний університет розвитку людини «Україна», Київ, Україна. https://orcid.org/0009-0009-7986-9203, anton89d@gmail.com*

Іван Даценко, *к.т.н., Національний університет Міністерства оборони України, Київ, Україна. https://orcid.org/0000-0002-0047-413X, docik_ivan@ukr.net*

Анотація. *У сучасному світі хмарних обчислень забезпечення безперебійної роботи й швидке відновлення після аварій є критично важливими аспектами для багатьох організацій. У статті розглядається використання AWS Global Accelerator як ефективного інструмента для забезпечення аварійного відновлення в хмарних мережах. AWS Global Accelerator дає змогу покращити доступність і продуктивність застосунків шляхом маршрутизації трафіку через глобальну мережу AWS, що забезпечує більш стабільне та надійне з'єднання.*

Дослідження охоплює аналіз основних функціональних можливостей AWS Global Accelerator, таких як підтримка високої доступності, зниження затримок і підвищення швидкості передачі даних. Особлива увага приділяється сценаріям аварійного відновлення, де Global Accelerator забезпечує швидке перенаправлення трафіку на резервні ресурси в разі збою основного компонента мережі. Це значно знижує час простою та мінімізує втрати даних, що є ключовим фактором для підтримки безперервності бізнес-процесів.

Також подано результати експериментальних досліджень, що демонструють ефективність використання AWS Global Accelerator у реальних умовах. Показано, що впровадження цього інструмента дає змогу знизити середній час відновлення після аварії на 30%, що значно покращує загальну стійкість хмарної інфраструктури. Наведено рекомендації щодо оптимального налаштування й використання AWS Global Accelerator для досягнення максимальних результатів у різних сценаріях аварійного відновлення.

Зроблено висновок про доцільність використання AWS Global Accelerator як невід'ємного компонента стратегій аварійного відновлення в хмарних мережах, що дає змогу забезпечити високу надійність та ефективність роботи сучасних інформаційних систем.

Ключові слова: *аварійне відновлення; хмарні системи; aws global accelerator.*

Introduction

Disaster recovery (DR) is a critical component of business continuity planning, designed to ensure that an organization can swiftly resume essential operations in the face of unexpected disruptions-whether due to natural disasters, cyberattacks, hardware failures, or human error. At the core of effective DR planning are two key metrics: RTO (Recovery Time Objective) and RPO (Recovery Point Objective).

RTO and RPO are two critical metrics used in disaster recovery planning to define the level of service the business requires from the IT infrastructure.

RTO, or Recovery Time Objective, is the target duration within which a system must be restored after a disruption in order to minimize the impact on business operations.

RPO, or Recovery Point Objective, is the maximum acceptable amount of data loss that a business can tolerate in the event of a failure. It defines how much data must be restored after a disruption and how frequently backups need to be performed to ensure the system can be recovered up to the latest point in time before the incident.

In this article, we talk about the application of AWS Global Accelerator as a key element in minimizing Recovery Time Objective (RTO) in disaster recovery. Through its effective utilization of the AWS global network, Global Accelerator directs user traffic via the most suitable AWS edge locations, thereby providing quicker and more stable connectivity to application endpoints—regardless of whether they are operating in active-active or active-passive disaster recovery architectures.

Through the utilization of sophisticated traffic management techniques and automated health checks, Global Accelerator is able to efficiently redirect requests from failed endpoints to functional ones in different AWS Regions or Availability Zones. This reduces downtime and enables recovery of services in seconds rather than taking minutes or hours.

Hence, the AWS Global Accelerator enhances performance and availability and also plays a key role in speeding up system failover in the event of outages being a valuable component of the disaster recovery toolkit of an organization.

Research

The system under investigation is a cloud-native, multi-region web application designed for high availability and fault tolerance. The architecture consists of services deployed across two AWS Regions (us-east-1 and us-west-2), fronted by a Global Accelerator endpoint, with Route 53, Elastic Load Balancing, and health checks integrated to support failover.

Key components include:

- **Primary and secondary AWS Regions** with mirrored infrastructure.
- **Amazon ECS** services running application containers.
- **AWS Global Accelerator** acting as a single-entry-point to route user traffic to the optimal healthy Region.

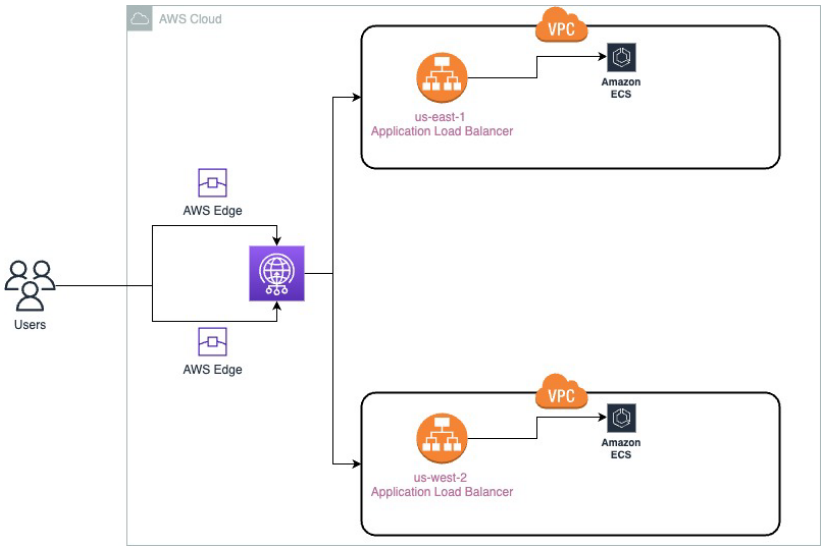

Fig. 1. Multi-regional application architecture diagram

AWS Global accelerator acts as a single point of entry, i.e. the CNAME for our software product is the Global Accelerator address. During the event of a failure in one region, it will automatically redirect traffic from users to another geographical region. Moreover, AWS automatically selects the region closest to the user and redirects traffic there, so the end-user will have the lowest latency to the application endpoint.

Global accelerator has been deployed with Terraform – a widely adopted Infrastructure as Code (IaC) tool. Terraform allows infrastructure components to be defined in a declarative configuration language (HCL), version-controlled alongside application code. This empowers teams to automate provisioning, track changes, and enforce standards across environments.

The deployed Global Accelerator was configured with:

- **Two endpoint groups**, each pointing to a Network Load Balancer in a different AWS Region.
- **Traffic dials** to simulate weighted traffic routing and failover logic.
- **Health checks** to detect unresponsive services and automatically redirect traffic away from impacted regions.
- **Cross-zone latency routing**, which allowed us to observe real-time behavior under failure.

```
endpoint_groups = {
  east = {
    endpoint_group_region        = "us-east-1"
    health_check_port            = 80
    health_check_protocol        = "HTTP"
    health_check_path            = "/"
    health_check_interval_seconds = var.health_check_interval_seconds
    threshold_count              = var.threshold_count
    traffic_dial_percentage      = 100
    endpoint_configuration = [{
      client_ip_preservation_enabled = true
      endpoint_id                  = data.aws_lb.alb-us-east-1.arn
    }]
    port_override = [{
      endpoint_port = 80
      listener_port = 80
    }]
  }
  west = {
    endpoint_group_region        = "us-west-2"
    health_check_port            = 80
    health_check_protocol        = "HTTP"
    health_check_path            = "/"
    health_check_interval_seconds = var.health_check_interval_seconds
    healthy_threshold_count      = var.threshold_count
    traffic_dial_percentage      = 100
    endpoint_configuration = [{
      client_ip_preservation_enabled = true
      endpoint_id                  = data.aws_lb.alb-us-west-2.arn
    }]
```

Here are the results of testing using the Apache Benchmark application which simulates user requests:

ab -n 100 https://URL

Where URL is Global Accelerator address, and respective load balancers addresses in different AWS regions.

Table 1

Global Accelerator testing results

Percentage of requests served within a certain time (ms)	us-east-1 region	us-west-2 region	Global Accelerator
50	387	270	269
66	393	290	277
75	406	326	291
80	411	354	335
90	495	480	403
95	1395	1262	528
98	2225	1604	1400
99	3411	2208	1456
100	3411	2209	1456

Results

Testing using Apache Benchmark showed significant improvements in query performance when using AWS Global Accelerator compared to directly accessing endpoints in the us-west-2 and us-east-1 regions. Key findings:

- **Reducing latency for average queries**

For 50% of requests, the service time through Global Accelerator was 269 ms, which is 118 ms less than us-west-2 and almost identical to us-east-1 (270 ms), indicating the efficiency of routing traffic through the nearest available points.

- **Stability under high loads**

In the range of 90-95% of requests, Global Accelerator exhibits significantly lower latency compared to regional endpoints. For 95% of requests, the service time through Global Accelerator was 528 ms, while for us-west-2 and us-east-1 these figures are 1395 ms and 1262 ms, respectively.

- **Maximum delays**

Global Accelerator provided significantly better performance at the extremes (98–100% of requests). The service time for 99% of requests was 1456 ms, which is significantly lower than the 3411 ms in us-west-2 and 2208 ms in us-east-1.

- **Route optimization efficiency**

Global Accelerator delivers superior performance by leveraging AWS's global infrastructure and automatic route optimization, reducing latency even under high loads and in geographically distributed environments.

Test results demonstrate that AWS Global Accelerator provides better performance and stability compared to direct access to endpoints in regions. This makes it an optimal solution for applications that require low latency, consistent performance, and high availability in global environments, but its use requires having a copy of the software product in multiple geographic regions, which increases the cost. Based on the research, this system was implemented in a real-world application deployed in AWS which helped to reduce downtimes during the geographical region outage.

References

1. Amazon Web Services. (n.d.). What is AWS Global Accelerator? AWS Documentation. [Electronic resource] / May 2025 – Mode of access: https://docs.aws.amazon.com/global-accelerator/latest/dg/what-is-global-accelerator.html

Дата першого надходження рукопису до видання: 22.05.2025
Дата прийнятого до друку рукопису після рецензування: 19.06.2025
Дата публікації: 25.07.2025