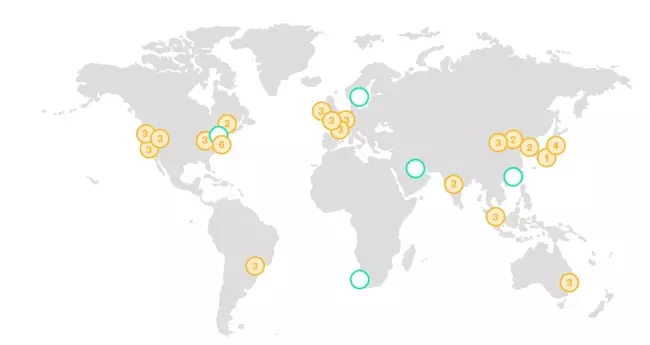
**AWS basic: regions, availability zones, VPC**

Khi mới bắt đầu tìm hiểu về AWS thì regions, availability zones, vpc là các khái niệm đầu tiên ta cần phải nắm được, ở bài dưới đây mình sẽ giải thích các khái niệm trên và xây dựng hệ thống thế nào cho phù hợp.



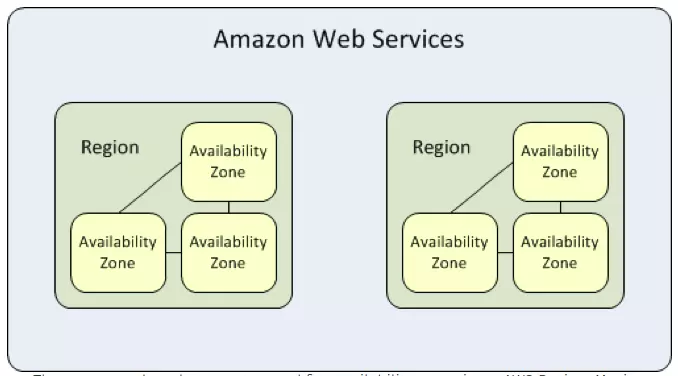
Vòng tròn màu vàng tương ứng với một region, con số ở trong thể hiện số availability zones có trong region đó, vòng tròn màu xánh là region sắp được xây dựng.

**1. AWS Region**

Các AWS region là các vị trí địa lí trên thế giới, nơi mà AWS tập trung xây dựng cơ sở hạ tầng của họ. Trong mỗi region chứa các khu vực khả dụng (availability zone - AZ), mỗi regin là độc lập với các regin khác về mặt cơ sở hạ tầng như khu vực, năng lượng, hệ thống cung cấp nước...Hiện nay AWS có khoảng 23 regions, region lớn nhất: us-east-1, có 5 availability zones

Sự cô lập này đảm bảo các yêu cầu về chủ quyền dữ liệu về việc dữ liệu người dùng không được rời khỏi một khu vực địa lý cự thể. Sự tổ chức các AWS Regin trên thê giới cũng rất quan trọng trong việc giảm độ trễ của hệ thống của bạn, tùy theo mật độ người dùng của chúng ta mà ta có thể chọn Region tại vị trí có nhiều người dùng nhất.

Trong mỗi Region sẽ có một hoặc nhiều vungf khả dụng (availability zone) với mỗi vùng là một trung tâm dữ liệu tách biệt với mỗi vùng khác, vậy thì tại sao lại có sự phân chia này thì mình sẽ giải thích ở phần sau.



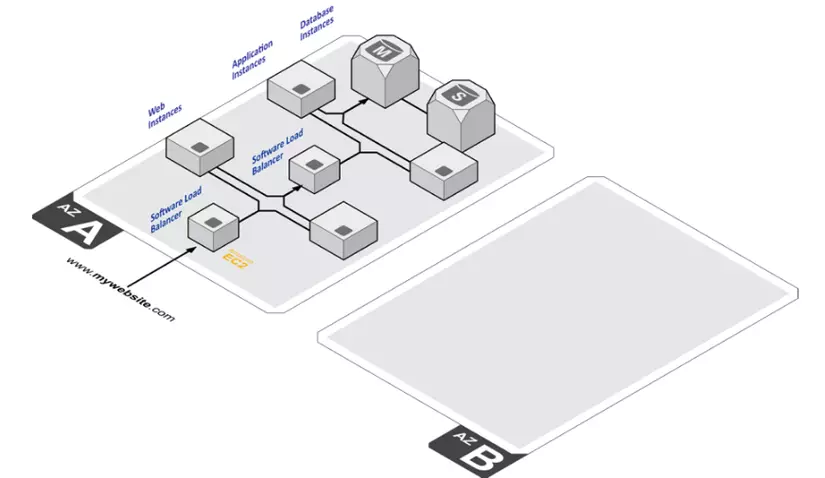
**2. AWS availability zones**

Một vùng khả dụng (availability zone) là một khu vực chứa trung tâm dữ liệu (data center). Mỗi availability zone có nguồn điện, mạng và kết nối dự phòng và riêng biệt để giảm khả năng hai khu vực bị lỗi đồng thời. Mỗi availability zone sẽ được hỗ trợ bởi nhiều data center vật lý, với availability zone lớn nhất được hỗ trợ bởi năm data center.

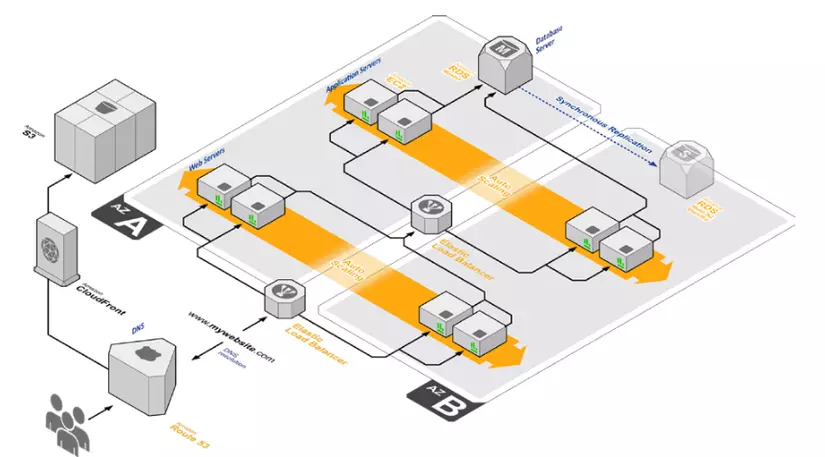
Một availability zone được mở rộng bởi nhiều data center vật lý, tuy nhiên các data center này là riêng biệt với mỗi availability zone, không có data center nào phục vụ đồng thời 2 hay nhiều availability zone.

Trong mỗi zone, các data center được kết nối với nhau bằng các đường dẫn mạng có độ trễ thấp. Tương tự, các zone trong một region giao tiếp với nhau thông qua các đường dẫn mạng riêng. Sự giao tiếp này nhăm mục đích đồng bộ, sao chép và phục hồi dữ liệu khi cần thiết của một số dịch vụ vủa AWS.

Vậy tại sao lại có sự phân chia nhiều availability zone trong một vùng? Biểu đồ dưới đây minh họa một một Region có 2 availability zone nhưng chỉ một zone được sử dụng để xây dự hệ thống. Kiến trúc phản ánh một mô hệ thống ứng dụng 3 tầng (web- app - DB). Với các máy chủ và cơ sở dữ liệu dự phòng được đặt chung trong một availability zone.

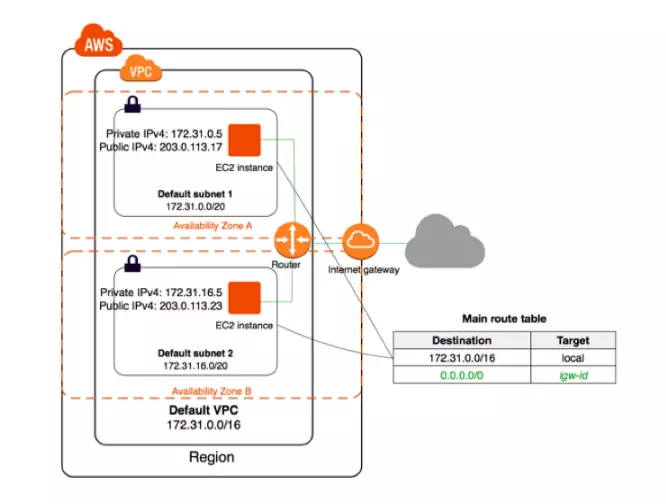


Tuy có các máy chủ dự phòng, kiến trúc hệ thống trên sẽ dễ phát sinh lỗi [single point of failure](https://en.wikipedia.org/wiki/Single_point_of_failure). Vì khi availability zone này gặp sự cố, toàn bộ hệ thống ứng dụng sẽ không hoạt động được nữa. Mô hình phân tán hệ thống ra nhiều zone dưới đây sẽ giải quyết được vấn đề này.



Bằng cách đặt các instance của mỗi lớp ở 2 zone riêng biệt, người dùng sẽ tránh được lỗi single point of failure. Bộ cân bằng tải (Amazon Elastic Load Balancers - ELBs) nằm ở các cấp ứng dụng khác nhau đảm bảo rằng ngay cả khi toàn bộ zone ngoại tuyến, lưu lượng truy cập sẽ được chuyển hướng đến zone thích hợp. ELB là một trong nhiều dịch vụ AWS có phạm vi Region và có thể trải dài qua các zone trong một region nhất định, vì thế nó vẫn hoạt động khi một zone gặp sự cố.

**3. AWS virtual private clouds**



VPC trong AWS rất rộng nên ở bài này chỉ nói về mặt khái niệm, chi tiết cụ thể sẽ giải thích và thực hành ở phần tiếp theo. VPC là một hệ thống mạng riêng ảo, phạm vi trong một AWS region, được sử dụng để:

* Cô lập các tài nguyên AWS của bạn khỏi các tài khoảng khác
* Điều hướng network ra và vào
* Bảo vệ các tài nguyên AWS của bạn khỏi các cuộc tấn công mạng

Bạn có thể đặt toàn bộ hệ thống vào một VPC để cô lập với môi trường bên ngoài. Bạn có toàn quyền kiểm soát môi trường mạng ảo của mình, bao gồm lựa chọn dải địa chỉ IP của riêng bạn, tạo mạng con và cấu hình bảng định tuyến và cổng mạng. Với VPC bạn có thể cấu hình để các thành phần nào có thể cho phép truy cập internet vào và ra (server, file store) và các thành phần nào không (database, ...)

**Lesson 3: Building For Resiliency**

Cam kết chất lượng dịch vụ AWS (SLA)

Resilient: phục hồi

**Lesson Recap**

* Server-based Services
* DynamoDB
* S3
* Compute Services

In this lesson, we looked at a variety of AWS data storage and compute options. AWS provides many options. Thinking through your use case, usage patterns, cost constraints and availability requirements are the key to successfully choosing the right services.

**Lesson Objectives**

You will be able to:

* Take advantage of different high availability options on AWS
* Create multi-AZ services
* Create multi-region services
* Identify what availability options exist for which AWS services
* Take advantage of resilient features in AWS services

**Glossary**

* **Active/Active:** A system that is running actively in multiple instances, typically in a distributed manner where complete functionality is available in more than one area.
* **Snapshot:** A complete copy of a dataset at a specific point in time.
* **Server-based Services:** Services that are existing applications that AWS provides as "managed services" and run on individual server instances.
* **DynamoDB:** AWS developed non-relational database
* **DynamoDB Global Tables:** Multi-Region DynamoDB Tables.
* **S3:** AWS developed object store that can store an unlimited amount of data.
* **Compute Services:** AWS services that provide generic compute capacity.

**Further Reading**

* [Redis](https://redis.io/)
* [ElasticSearch](https://www.elastic.co/)
* [MongoDB](https://www.mongodb.com/)
* [DynamoDB Pricing Models](https://aws.amazon.com/dynamodb/pricing/)
* [Amazon Prime Day 2019](https://aws.amazon.com/blogs/aws/amazon-prime-day-2019-powered-by-aws/)
* [S3 Pricing](https://aws.amazon.com/s3/pricing/)
* [WannaCry randsomware attack](https://en.wikipedia.org/wiki/WannaCry_ransomware_attack)
* [CloudFormation](https://aws.amazon.com/cloudformation/)
* [Terraform](https://www.terraform.io/)

**Lesson 4: Business Objectives  
Lesson Outline**

* Uptime
* Downtime
* RTO
* RPO
* Disaster Recovery

**Lesson Objectives**

You will be able to:

* Calculate availability in terms of up and down time
* Set reasonable business metrics for RTO and RPO
* Make determinations on what types of DR plans a company needs
* Implement a DR plan

<https://wikimaytinh.com/thoi-gian-uptime-la-gi.html#:~:text=Uptime%20c%C3%B3%20ngh%C4%A9a%20l%C3%A0%20th%E1%BB%9Di,l%C3%AAn%20t%E1%BB%9Bi%2099%2C9%25>.

**Uptime là gì?**

Uptime là gì? Uptime có nghĩa là thời gian hoạt động. Uptime đề cập đến lượng thời gian máy chủ được thiết lập và chạy. Khi tìm kiếm một nhà cung cấp dịch vụ lưu trữ, bạn sẽ thấy rằng họ đề cập rất nhiều đến thời gian hoạt động Uptime lên tới 99,9%.

**Downtime là gì?**  
Downtime có nghĩa là thời gian gián đoạn dịch vụ (Interruptions to service – tức là khi trang web của bạn không khả dụng) được gọi là thời gian chết. Nhà cung cấp đảm bảo Up

**Nhà cung cấp đảm bảo Uptime là gì?**

Một số cung cấp đảm bảo thời gian hoạt động. Điều này có nghĩa là nếu thời gian hoạt động của họ giảm xuống dưới một mức nhất định (thường là 99,9%), họ sẽ cung cấp cho bạn tín dụng cho thời gian ngừng hoạt động bổ sung như một hình thức đền bù. Tìm các công ty lưu trữ có mức Uptime cao nhất: bạn muốn trang web của mình luôn sẵn sàng khi khách truy cập.

**RTO – Recovery Time Object**

RTO (Recovery Time Object), dịch nôm na là thời gian phục hồi tính từ thời điểm chủ doanh nghiệp đưa ra yêu cầu restore lại RPO chỉ định.

Ví dụ như phòng nhân sự mất file tính lương và họ yêu cầu bạn trong vòng 1 tiếng phải phục hồi cho họ ngay, như vậy RTO = 1 giờ.

*Cũng như RPO, RTO cũng đóng vai trò tỷ lệ nghịch với chi phí, RTO càng nhỏ thì chí phí càng cao.*

**RPO – Recovery Point Object**

RPO là viết tắt của Recovery Point Object hay còn gọi là ” thời điểm phục hồi”.

Thời điểm ở đây là thời điểm cụ thể, ví dụ như 1 phút trước, 1 ngày trước, 1 tuần trước hoặc 1 tháng trước, đạ khái là 1 thời điểm đã xảy ra và doanh nghiệp muốn dữ liệu được phục hồi vào đúng thời điểm mà họ mong muốn.

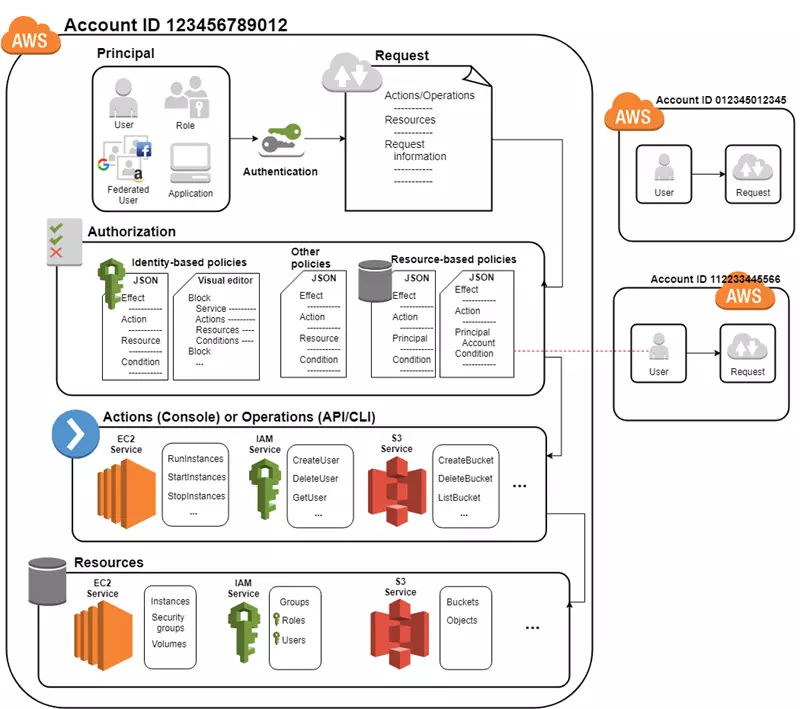
Ví dụ như: Công ty ABC muốn hệ thống backup của họ có khả năng phục hồi dữ liệu 1 ngày trở về trước từ thời điểm cần phục hồi dữ liệu thì RTO là 1 ngày. Hay một công ty khác muốn có thể khôi phục lại dữ liệu ở bất kỳ thời điểm nào thì RTO của họ là zero.

**Uptime Calculation**

Allowed downtime = (30 days 24 hours 60 minutes) - (30 days 24 hours 60 minutes \* SLA percentage[Cam kết chất lượng dịch vụ AWS])

Công việc tuyệt vời! Vì đây là ứng dụng quốc tế, công ty có thể không xác định được bất kỳ múi giờ cụ thể nào mà mức sử dụng là tối thiểu. Tương tự, ngày đầu tiên của tháng có thể là bất kỳ ngày cụ thể nào trong tuần. Tuy nhiên, ***Thứ Bảy là những ngày có lưu lượng truy cập thấp đối với các trang web kinh doanh trên toàn thế giới***. Tất nhiên, một số trang web nhất định có thể là ngoại lệ, vì vậy công ty nên xem xét các mẫu lưu lượng truy cập cụ thể của họ trước khi chọn thời hạn bảo trì.

https://calculator.aws/#/

**IAM (identity and access management)**https://viblo.asia/p/tong-quan-phan-quyen-tren-aws-identity-and-access-management-m68Z0Aj6lkG  


**Glossary**

**Authentication**

Who can sign in and use the API.

**Authorization**

What permissions that user has.

**IAM Users**

An IAM User is a resource that represents a person or application and allows that person to interact with AWS. It consists of a name, console login credentials, and API keys.

**Instance Profile Roles**

Instance profile role is a special role that is assigned to EC2 instances which allow application running on those instances to obtain temporary credentials aligned with that role.

**Assume Role**

When I say a server can "assume a role", I mean the application can obtain temporary credentials that are aligned to a role from AWS.

**Root User**

The root user of the AWS account is the email address that was used to create the account. This user has full control of the account and its credentials should be treated with the highest sensitivity

**Multi-Factor Authentication (MFA)**

An authentication mechanism in which a user must present two or more peices of evidence (or factors) to an authentication system. In AWS the user can setup multi factor authentication by setting up a virtual or physical MFA device which they would need to physically posses and present in addition to their password for logging in.

**RBAC**

Role Based Access Control refers to the model of allowing users permission to use roles which they have been assigned to as part of their job function. Specific roles would provide access and permissions to specific systems and actions.

**Identity Federation**

Identity Federation enables you to manage access to your AWS resources centrally. With federation, you can use single sign-on (SSO) to access your AWS accounts using credentials from your corporate directory. Federation uses open standards, such as Security Assertion Markup Language 2.0 (SAML), to exchange identity and security information between an identity provider (IdP) and an application.

**IAM Roles**

AWS IAM identity in an AWS account which provides specific AWS permissions. Authorized entities, such as IAM users, identity providers, and AWS services shall assume a role and gain the permissions assigned with that role. No permanent password or credentials exist with an IAM role. Users who assume the role will have access to temporary credentials to make API calls.

**IAM Policies**

Document(s) assigned to IAM users and roles that specify what actions that user can take in AWS, whether to allow or deny the action, what specific resources that action can be executed on, and any other conditions related to the permissions.

**Resource Policies**

Policy documents associated with a specific resource in AWS, such as an S3 bucket, which defines who can perform what actions, with specific resources and conditions as needed. Similar to an IAM policy except it is associated with a resource and the actor or principal needs to be defined.

**Attribute Based Access Control**

Attribute-based access control (ABAC) is an authorization strategy that defines permissions based on AWS tags. Tags can be attached to IAM principals (users or roles) and to AWS resources. You can create a single ABAC policy or small set of policies for your IAM principals. These ABAC policies can be designed to allow operations when the principal's tag matches the resource tag.

**Least Privilege Access**

When creating IAM policies that provide permissions to users and roles, we want to follow the common security practice of granting least privilege. Least Privilege means we grant only the permissions required to perform the necessary tasks.

**Control Plane**

Also known as the **management plane**, this is the ability to interact with AWS in order to provision and configure services.

PREVIOUSNEXT