Tài liệu yêu cầu phần cứng, phần mềm đáp ứng HA

1. Giới thiệu
2. Giới thiệu về High Availability

High availability có nghĩa “Độ sẵn sàng cao”. Tức là những máy chủ hoặc thiết bị luôn trong tình trạng sẵn sàng phục vụ, giảm thiểu khả năng gián đoạn của hệ thống. Hoặc có thể hiểu đơn giản High availability là một giải pháp hoặc quy trình hay công nghệ nhằm thực hiện chức năng đảm bảo cho ứng dụng, cơ sở dữ liệu có thể truy cập được 24/7 trong mọi điều kiện. High availability có nghĩa “Độ sẵn sàng cao”. Tức là những máy chủ hoặc thiết bị luôn trong tình trạng sẵn sàng phục vụ, giảm thiểu khả năng gián đoạn của hệ thống. Hoặc có thể hiểu đơn giản High availability là một giải pháp hoặc quy trình hay công nghệ nhằm thực hiện chức năng đảm bảo cho ứng dụng, cơ sở dữ liệu có thể truy cập được 24/7 trong mọi điều kiện.

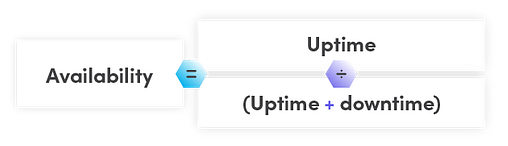
1. Giới thiệu về SLA

SLA (Service-Level Agreement) là thỏa thuận giữa nhà cung cấp và người dùng về các chỉ số có thể đo lường như thời gian phục vụ, khả năng đáp ứng và trách nhiệm.

Các metrics được sử dụng để cam kết trong SLA:

1. Availability

* Thời gian “sống”, uptime của hệ thống, đo trong một khoảng thời gian



* Thường được tính theo số số 9:
  + VD: 2 số 9: 99%, 3 số 9: 99,9%, 4 số 9: 99,99%
  + Với SLA 99%, hệ thống chạy 24/7 cần cam kết thời gian downtime không quá:
    - 14m24s/ngày
    - 1h30m48s/tuần
    - 7h14m41s/tháng
    - 3d14h56m18s/năm

1. Mean time to recovery

* Thời gian trung bình cần để xử lý sự cố

1. Response time and resolution time

* Thời gian ghi nhận báo cáo sự cố và thời gian giải quyết

1. Error rates

* Tỉ lệ xảy ra sự cố trong một khoảng thời gian

1. Security and compliance

* Khả năng bảo mật của hệ thống

1. Các chiến lược để đạt được HA
2. Replication

* Nhân bản các thành phần của hệ thống để đảm bảo nếu có sự cố xảy ra tại một instance thì instances khác sẽ tiếp tục thế chỗ xử lý yêu cầu

1. Load balancing

* Chia workloads ra nhiều node/server khác nhau, tang tốc độ xử lý và tránh quả tải

1. Failover Clustering

* Đảm bảo HA bằng cách tạo cụm các server. Nếu một server chết, server khác sẽ thay thế.
* Thường được sử dụng trong hệ thống cơ sở dữ liệu

1. Distributed Data Storage

* Lưu dữ liệu replica tại nhiều nơi, đảm bảo khả năng khôi phục khi gặp lỗi

1. Monitoring, Log tracing and Alert

* Theo dõi tài nguyên hệ thống
* Thu thập log
* Cảnh báo sớm các lỗi có khả năng xảy ra

1. Đề xuất triển khai
2. Môi trường Server
   1. Cơ sở dữ liệu:

Triển khai cụm PostgreSQL sử dụng stack HAProxy, Keepalived, Patroni và Etcd để đảm bảo tính sẵn sang cao cũng như tăng hiệu suất truy vấn