Apunte - Semana 13 - 21/10/2022

Max Richard Lee Chung - 2019185076

> La prueba corta 5 y 6 se entrega el sábado 22-10 a más tardar a las 11:59 a.m.

Bases de datos en la nube

Red privada es como red de la casa como un internet provider contratado de los routers y se conecta a las computadoras.

Dentro de la nuve, la red privada se llama VPC (Virtual Private Cloud), el cual asigna un rango de IP's que se asigna una IP a una computadora cuando se conecta a alguna red. Por defecto se usa la 10.0.0.0/8 y se puede usar un subnet calculator para calcular la cantidad de host que se pueden usar, como por ejemplo, 16 millones de computadoras.

Existen dos formas de organización, las zonas privadas y las zonas públicas (DMZ o desmilitarizada). Dentro de las zonas privadas, viven todos los datos de los recursos más importantes de una empresa o compañía y normalmente contienen el backend (base de datos). Dentro de las zonas públicas, viven todos los datos que se exponen en internet y normalmente contienen el frontend (API's, websites, entre otros) dado que es donde los clientes van a interactuar. Sin embargo, si el API no es consumido por clientes, se colocal dentro de la zona privada.

Se selecciona una región (zona geográfica) para crear una red privada de bases de datos, como por ejemplo AWS Region. Hay regiones que están diseñados o especiales para mantener datos especiales. La región define dónde se va a localizar el servicio guardado. Las zonas de disponibilidad son datacenters físicos o collocations que son edificios que siguen muchas normas para mantener normas computacionales. Hay lugares que tienen varios datacenters separados para poder recuperar los datos en casos de desastres, ya que si se encuentran muy cercanos, pueden afectar todos los datacenter y por lo tanto, la disponibilidad.

Una vez que se selecciona la región, se seleccionan subredes (al menos 3) en ambas zonas de organización. Se crean al menos 3 por los algoritmos de consenso para seleccionar el maestro y los esclavos para delegar trabajos y delegar el rol de maestro a algún esclavo si el maestro principal falla. Si en una zona de disponibilidad, un maestro y esclavo falle, la aplicación se encontrará fuera de servicio. Se debe de planificar los recursos y tolerancia del tiempo fuera de servicio para garantizar que la aplicación sea aceptada y funcional.

AWS Services

N. Virginia es la zona más estable y vieja. A veces las regiones más nuevas, tienen servicios más nuevos. Las zonas privadas se desactiva el "auto-assign public Pv4 adress" para garantizar que cualquier servicio creado en la red, no pueda ser accesible desde internet. Las zonas privadas sí tiene habilitada tal opción para garantizar que se pueda observar desde internet.

RDS es un servicio para crear diferentes bases de datos. Dependiendo del caso de uso o cargas de trabajos, se define qué tanto se quiere personalizar para optimizarlo dentro del caso, dado que se crean para que sirvan en muchos casos de usos. Dentro de las credenciales, se puede almacenar como los secrets de los helm charts. Se puede configurar el tipo de estructura de organización (estructura de índices) en la base de datos.

El network de la máquina es muy importante para poder responder de forma rápido las actualizaciones, como por ejemplo, una consistencia fuerte, es preferible que los cambios se realicen en el menor tiempo posible. Al buscar sobre una máquina, un núcleo es un procesador físico que permite un hilo (thread). Dicha información es importante para permitir trabajar simultaneamente las consultas y reducir el tiempo de respuesta. La memoria afecta al memory footprint, dado que entre más memoria, más datos se pueden tener para no tener que ir a disco, sin embargo, el costo aumenta por la calidad de memoria.

Discos en red tiene que conectarse a las máquinas a través de una red, es decir, los discos no están conectados a las máquinas. Los discos físicos son muy rápidos en comparación de las de red, ya que están limitados a la velocidad de la tarjeta de la red y también por el internet que están conectados la base de datos y el cliente. Si se utilizan discos EBS de la nube, quiere decir que pierde un 40% de rendimiento en comparación de un disco de estado sólido ya que hay que estar moviendo datos en red. El ancho de banda quiere decir qué tan rápido se puede acceder o transportar información. El ancho de banda es una red exclusiva sin control de almacenamiento que va a estar conectada a cualquier máquina creada ya que todo tipo de máquina ocupa disco en red. El rendimiento de red es qué tan rápido es la red que existe entre los nodos de la base de datos. Hay tipos de máquinas que soportan discos de estados sólidos y discos EBS, sin embargo, los precios pueden aumentar demasiado ya que pueden ser mucho más rápidos.

> Uno siempre se debe hacer análisis de rendimientos y costos de nuevos servicios para el caso de uso indicado.

El tipo de almacenamiento se puede analizar por el monitoreo de entradas y salidas de datos. Cada giga de tamaño en el disco, aporta una gran cantidad de operaciones de entrada y salida, sin embargo, se puede configurar los IOPS (operaciones de entradas y salidas). Multi-AZ deployments son varias instancias de stand-by y slave para tener redundancias de datos y réplicas.

Se pueden crear grupos de subredes para seleccionar cuál subred va a utilizar un datacenter para que los datos estén más cerca a los usuarios (evitar los saltos entre router para disminuir el viaje de los datos entre ellos y mejora el rendimiento). Se puede crear un grupo de seguridad para ver qué datos están cerrados o abiertos.

> Los backbones son los que tienen las mejores tecnologías ubicados en países desarrollados.

Identity and Access Management (IAM) es una definición de usuarios para definir un rol (conjunto de permisos que tiene el usuario) que están conectados a políticas. Se puede usar un modelo de múltiples capas de seguridad para aumentar la dificultad de robos de datos. La aplicación puede tener un rol para verificar la contraseña y acceso de dato.

La encriptación podría deshabilitar la posibilidad de robo de datos al entrar a la subred privada, ingresar a los datos de la base de datos y copiar toda la información dado que la información encriptada no se puede leer de manera fácil. Sin embargo, con respecto al rendimiento de bases de datos, no debería de existir la encripción ya que recuperar los datos implica desencriptar los datos y cada escritura tiene que volver a encriptar la información.