cloudwatch

- servicio de monitorización
- la mayoría de servicios incorporan de base algún sistema de monitorización de manera automáticamente
- se pueden personalizar ($)
- High Resolution Custom Metrics
    - permite métricas cada segundo
    - la retención de métrica pasa de 14 días a 15 meses
    - las métricas de menos de 60 segundos están disponibles 3 horas
    - de más de 60 segundos, están disponibles 15 días
    - de más de 300 segundos, están disponibles 63 días
    - de más de 3600 segundos, están disponibles 455 días
    - al pasar su periodo de disponibilidad, se agrupan, perdiendo detalle
- despliegue por regiones
- namespaces (se agrupa por servicios)

EC2

- CPU
- Network
- Disk
- Status Check
    - System Status Check : hard por debajo del server (necesario STOP y START para que cambie de hard. No sirve reboot)
    - Instance Status Check : la VM falla, Reboot (mala configuración network, sin memoria, problemas de disco..)
    - health chechs van cada minuto
- 5 minutos de intervalo
- Custom metric : RAM, espacio disco usado/libre
- (mirarse la métricas disponibles)
- SSD
- HDD (NO ROOT volumen)
- … mirar números

Alarmas

- Crear alarmas
    - permite notificaciones
    - permite tomar acciones

I/O Credits

- credito inicial : 5.4 millones de I/O credits, permite 3000 IOPS durante 30 minutos (pensado para el arranque y sobretodo en volúmenes pequeños)
- los discos de más de 1024GB tienen una performance igual o superior a estos 3000 IOPS que nos ofrecen en créditos, este crédito no computa
- 3 IOPS/1GB
    - ejemplo: 100GiB tiene 300IOPS
    - una manera de mejorar el rendimiento se puede subir la capacidad del disco (que hace que aumenten los IOS)

    - o aumentarlos desde el manejo de EBS - volumen
        - también el tipo de disco
    - en ambos casos, $

EBS Volumes

- para conseguir el rendimiento óptimo
- inicialización : cuando recuperas un snapshot desde un S3, ejecutar un comando (dd de lectura) para optimizarlo
- Métricas EBS:
    - VolumeQueueLength : cantidad de peticiones en cola pendientes
- Status:
    - OK
    - warning : degradado o bastante degradado
    - impaired : no disponible o totalmente degradado.

RDS

- database connection
- Disk Queue Depth
- Free Storage Space
- ReplicaLag
- Rea/Write IOPS
- Read/Write Latency

ELB

- 60 segundos por defecto
- 3 tipos:
    - classic (no recomiendan su suo)
    - capa 7
    - capa 4
- métricas:
    - SpilloverCount : rechazadas
    - SurgeQueueLength : cola
    - HTTPCode_ELB_5XX_Count
    - RequestCountPerTarget
    - TargetResponseTime

ElasticCache

- redis / memcached
- Métricas
    - CPUUtilization
        - memcached:más allá del 90%, se ha de escalar (más nodos)
        - Redis: en función del número de cores
    - SwapUsage : mala señal si se usa SWAP en un servicio
        - por encima de 0 es que algo va mal, pero sobre todo por encima de 50MB → aumentar ConnectionOverhead
    - Evictions : falta de memoria

        - reemplazo de items nuevos por viejos
    - CurrConnections
        - en función de la aplicación desplegada, hay que «jugar» con los valores para montarnos nuestras alarmas

Consolidated Billing - AWS Organizartions

- descuento por volumen (cuanto más servicios consumes, más baratos te salen)
- tengo una cuenta «madre» y enlazo las cuentas hijas para no perder el volumen
    - en la madre no debería haber recursos desplegados
    - uso de tags
- instancias reservadas EC2 (del mismo tipo, región, AZ) compartidas entre las diferentes cuentas
    - market place : venta de EC2 reservadas (si ya no la necesitamos)
- AWS Organizartions
    - creación de unidades organizativas
    - gestión de cuentas hijas (estructura) ≡ ldap
    - billing alerts

CloudTrail

- activado por región por cada AWS Account
- auditoria
- puedes consolidar los logs en un bucket S3
    - con una bucket policy para permitir acceso cruzado de cuentas

Elasticity and Scalability

- elasticidad : capacidad de ajustar las necesidades en función de la demanda (para no pagar más de lo necesario ni quedarse corto)
    - basada en el tiempo : previsión en función de horarios/días
    - basada en el volumen
    - combinando monitorización, tagging y automatización, se puede optimizar todo
    - más instantáneo
- escalabilidad:
    - incrementar la carga de trabajo sin que se vea afectada la «performance»
    - posibilidad de expandir sin limitación en el tiempo
        - vertical : scale-up : autmentar tamaño del recurso
            - no es elasticidad
            - por ejemplo, cambiar el tipo de instancia de un EC2 por necesidad
            - tiene un techo
        - horizontal : scale-out : aumentar el número de recursos
    - más a largo plazo

AWS High Availability & Fault Tolerance

- Fault Tolerance : continuar funcionando aunque haya un fallo de hardware
    - servicios que por si solos están en estas categorias:
        - S3
        - AuroraDB
        - SQS

        - ELB
    - servicios que hay que «configurar» o añadir algún servicio adicional para que funcionen así
        - EC2
        - EBS
        - montar en diferentes AZ, Elastic IP
- Regiones, múltiples AZ
    - cada AZ es independiente a errores
    - todo el tráfico de 1 región es de baja latencia y sin coste
- soluciones multi-site

RDS y multi AZ

- el multi AZ en RDS crea una réplica (no accesible) sincronizada por si la primera cae y hace el cambio de endpoint en ese caso
- replicación síncrona
- los backups o parches se aplican en la «réplica», para evitar perdida de performance en la principal
- usar provisoned IOPS para consitencia
- Read Replicas
    - Aurora, MySQl, MariaDB, PostgreSQL
    - asíncrona
    - accesibles
    - gran carga de trabajo de lectura
    - se han de borrar especificamente (si borras la primeria, una de las Read Replicas se promociona)
    - Cross-Region Replication : permite crear read replicas en otras regiones
    - se permiten crear read-replicas de read-replicas
- Aurora:
    - más chupi
    - por defecto trabaja / se «replica» en 3 AZs (sin «usar» el MultiAZ de las otras)

Bastion Hosts

- subnet pública
- conectar desde internet para gestionar tu infraestructura
- actua de proxy entre tus servidores e internet, permitiendo acceso a tu infraestructura

AutoScaling problemas

- no existe el security group
- no existen las claves
- la configuración no está soportada
- no existe el grupo de autoescalado
- ...

altran

- Direct Connect

From:
https://matebcn.eu/ - **miguel angel torres egea**

Permanent link:
https://matebcn.eu/info:cursos:itformacion:awsassociate:cloudwatch

Last update: **22/10/2018 20:09**