Filipe Borba e Gabriel Moreira

R0 - Introdução

Explorando a AWS

1) Descreva passo a passo o processo utilizado:

1- Fizemos login na AWS com o usuário e senha fornecidos.

2- Procuramos na aba Services qual era o serviço que administrava os usuários. Chegamos no grupo de "Security, Identity & Compliance" no serviço IAM.

3- No menu lateral, embaixo de Dashboard, criamos um novo grupo na aba Groups.

4- Demos o nome do grupo de Grupo4, clicamos em Next Step e em seguida, escolhemos as duas primeiras opções de permissões: AdministratorAccess e IAMFullAccess. Por fim, revisamos se tudo estava correto e criamos o grupo.

5- Com o grupo criado, voltamos para o Dashboard e entramos na aba Users, clicando no botão "Add user".

6- Colocamos dois usuários com a ajuda do botão "Add another user" sendo eles "filipefb" e "gabrielfm". Em seguida, selecionamos os dois tipos de acesso: "programmatic access e AWS Management Console access", mantendo os campos Console password (autogenerated password) e Require password reset.

7- Selecionamos as permissões do Grupo4 para os dois usuários e criamos os dois usuários.

2) O que são Policies? Por que elas são importantes e devem ser bem definidas?

Policies são entidades que definem as permissões de um determinado usuário ou grupo a respeito de um recurso na AWS. Elas são importantes e devem ser bem definidas porque, caso contrário, pode causar uma confusão, como por exemplo: um usuário pode acessar coisas que não devia ou um administrador pode não ter o poder de controlar sua aplicação.

Primeira Instância

1) Descreva passo a passo o processo utilizado:

1- Fizemos login na AWS com o usuário e senha fornecidos.

2- Em Build Solution clicamos em Launch a virtual machine.

3- Na próxima página selecionamos a opção "Ubuntu Server 16.04 LTS (HVM), SSD Volume Type".

4- No passo 2, selecionamos a opção "t2.micro" como nosso tipo de instância.
5- No passo 3, nada foi mudado, com isso pulamos diretamente para o passo 4.
6- No passo 4 (Add Storage), mudamos o "Size (GiB)" para 8.
7- No passo 5, nada foi mudado, com isso pulamos diretamente para o passo 6.
8- No passo 6, em "Assign a security group", selecionamos a opção "Create a new security group" - "copy to new",
Feito isso, trocamos o "Source" de "custom" para "my IP".
9- Escolhemos a primeira opção no campo "Security Group ID", fizemos uma revisão de tudo e clicamos em "Launch".
10- Criamos uma nova "key pair" e fizemos o download. Em seguida, lançamos a instância.

2) Dentro do contexto de Cloud Computing, defina os termos: instance, image, region, VPC, subnet, securitygroup.

Instance: é uma máquina virtual disponibilizada por um serviço privado ou público de Cloud.
Image: são imagens de disco pré instaladas que são customizadas para serem usadas em Clouds públicas ou privadas.
Region: são as áreas onde os recursos da Cloud pública estão fisicamente instaladas. Uma escolha de uma determinada região em detrimento de outra pode corrigir problemas como latência.
VPC: significa Virtual Private Cloud e é uma ferramenta que os serviços de Cloud normalmente disponibilizam para o usuário definir sua própria rede virtual com a possibilidade da escalabilidade.
Subnet: é uma divisão da rede IP em menores segmentos de redes, melhorando sua eficiência.
Security Group: é um firewall virtual que controla o tráfego tanto de entrada quanto de saída. O Security Group cuida das permissões e da segurança da instância.

3) O poder computacional das instâncias é medido em vCPU. O que é vCPU?

Uma vCPU é uma CPU física que é designada à uma máquina virtual. Uma ou mais vCPUs são designadas à cada máquina virtual em um ambiente de cloud, contudo, a máquina virtual entende que cada vCPU é apenas um core da CPU física. Se a máquina hospedeira tiver múltiplo múltiplos cores disponíveis, então a vCPU acaba sendo um espaço de tempo de processamento entre todos os cores disponíveis, permitindo que múltiplas máquinas virtuais sejam hosteadas em um número menor de CPUs fisicas.

Primeiro Deploy - Ghost Blog Platform

1) Quantas instâncias foram criadas automaticamente? Criar instâncias automaticamente é um atributo
positivo ou negativo?

Foram criadas 4 instâncias em máquinas diferentes. A criação de instâncias automaticamente é um atributo tanto negativo quanto positivo. Do lado positivo, pode-se analisar que é um instrumento que poupa muito trabalho caso a criação de múltiplas instâncias fosse necessária. Por outro lado, é uma ferramenta que pode ser devastadora se algum erro fosse cometido, ou operado por pessoas com más intenções, acarretando em um enorme pagamento pelas instâncias caso muitas fossem criadas desnecessariamente.

2) Quanto custou o protótipo? Assuma que usou uma hora de cada instância.

Utilizando a ferramenta "https://calculator.s3.amazonaws.com/index.html", concluímos que as 5 instâncias criadas custariam $0,26/hora.

Limpando a bagunça (força bruta)

1) Dada a quantidade de computadores apontada na questão anterior, descreva como você montaria um ambiente Ghost em um Datacenter próprio. Assuma que você ainda não possui nenhum hardware disponível, apenas um orçamento aprovado.

Em um Datacenter próprio, poderíamos colocar as 2 instâncias da t2.large em dois computadores poderosos e as 3 instâncias da t2.micro e t2.medium em outros computadores mais modestos. Com isso, teríamos 4 a 5 máquinas dedicadas para o ambiente Ghost, divididas por responsabilidade, não necessitando de mais recursos. Essa divisão por responsabilidade é interessante, pois podemos garantir que se houver alguma falha de hardware, apenas uma parte da aplicação ficará fora do ar, e não ela inteira. Além disso, teríamos que nos preocupar com o gasto de energia mensal e com a manutenção do equipamento no caso do Datacenter próprio.

2) Agora, somando o fato de que hardware deprecia com o tempo e possui um custo mensal de manutenção, compare em termos de custo uma Public Cloud e um Datacenter próprio.

Considerando o cenário apresentado, o Datacenter próprio compensaria seu custo em detrimento do Public Cloud depois de aproximadamente 1 ano, dado os custos de R$688 por mês no Public Cloud e R$7000 com equipamentos no começo e uns R$100 de gastos fixos, como energia, para o Datacenter próprio. No entanto existe todos os fatores de depreciação do hardware, espaço, gasto energético, escalabilidade, etc. que acabam dificultando a manutenção de um Datacenter próprio.

A construção do Datacenter próprio acaba sendo muito trabalhosa e custosa num primeiro momento, portanto, é interessante adotar a utilização de um Public Cloud no começo para facilitar a questão de infraestrutura e escalabilidade.

Escalabilidade

1) O que faz o crontab?

O crontab é apenas um editor do arquivo que serve de base para o programa cron. Nesse arquivo, são especificados o horário, dia, período e frequência que um determinado comando irá ser executado. O cron, por sua vez, cuida da execução
de tais comandos na hora correta.

Montando o Autoscaling Group

Endereço: http://lbr0escalabilidade-929514452.us-east-1.elb.amazonaws.com/

1) O que é uma AMI?

Uma Amazon Machine Image (AMI) é basicamente uma cópia de uma instância existente. Ela inclui as especificações (Sistema operacional, aplicativos, etc.), as permissões de execução e um mapeamento dos volumes que precisam ser anexados à instância.

2) O que faz o LoadBalancer? Explique o algoritmo utilizado.

O LoadBalancer distribui o tráfego da aplicação em várias instâncias a fim de evitar o congestionamento de tais endereços. A condição para um usuário ser redirecionado para um determinado endereço depende da regra criada pelo administrador e sua prioridade. Se

um endereço tem as condições atendidas e sua prioridade é maior, o usuário será encaminhado para o mesmo.

Fazendo uso da Escalabilidade Horizontal

1) Enfim, como funciona o Autoscaling Group da AWS?

O Auto Scaling da AWS monitora as instâncias ativas e ajusta automaticamente a capacidade da aplicação para que a demanda seja sempre suprida e com somente o que for necessário, reduzindo o custo de ter, por exemplo, máquinas ociosas. O Auto Scaling Group nada mais é do que uma coleção das instâncias do EC2 que são necessárias para o serviço.

2) Qual a diferença entre escalabilidade horizontal e escalabilidade vertical?

Associamos a escalabilidade vertical ao processo de adicionar recursos à um sistema já existente, enquanto que a escalabilidade horizontal visa distribuir a demanda em mais sistemas. Por exemplo, um computador poderoso dá conta de um servidor de uma empresa. Com o crescimento dela, porém, as requisições aumentam e o computador começa a não dar conta de processá-las. Para escalar verticalmente esse sistema, compraríamos um processador melhor, mais RAM, um SSD com mais capacidade. Já para escalar horizontalmente esse sistema, compramos mais máquinas iguais ou piores que a original, possibilitando a distribuição das requisições.

3) Qualquer serviço pode fazer uso desse modelo?

Para o caso de um serviço de banco de dados, este modelo não seria válido, uma vez que seria necessário sincronizar cada ação para as outras instâncias. Por conta disso, a abordagem mais correta seria escalar verticalmente este sistema. Apesar disso, existe uma grande variedade de serviços que podem usufruir desse modelo, porém não é recomendado, uma vez que estar preparado para escalabilidade horizontal sem necessidade acaba saindo mais caro do que o normal. É necessário verificar as necessidades do serviço antes de implementar essa solução.

Questões Complementares

1) O que é um VPS? Qual a diferença entre uma instância e um VPS?

Um VPS (Virtual Private Server) é um servidor dedicado fragmentado em vários outros pequenos servidores, para ser melhor aproveitado pelos usuários. É um serviço que possui mais e melhores recursos como memória e processador. A instância, por sua vez, é conjunto de máquinas virtuais que operam em diferentes hardwares. A principal diferença é a escalabilidade que uma instância oferece, uma vez que VPS possui um hardware limitado.

2) Defina Platform as a Service (PaaS) e Software as a Service (SaaS).

PaaS: É uma plataforma de gerenciamento e desenvolvimento de aplicações inteiramente na nuvem. O usuário dessa plataforma não deve se preocupar com sua estrutura e configurações como os servidores e o sistema operacional, uma vez que fazem parte do serviço.

SaaS: É uma solução de software oferecido por alguma empresa inteiramente hospedado na nuvem. O usuário não deve se preocupar com nenhuma configuração, nem mesmo do próprio software. Um exemplo de SaaS é o pacote Office 365

3) O modelo LoadBalancer possui um custo fixo elevado. Como você justificaria o uso e a configuração de um LoadBalancer para uma empresa?

Podemos justificar o uso de um LoadBalancer para uma empresa quando verificamos que há uma alta variação da demanda de seus recursos computacionais. Pensando apenas no Brasil, numa empresa como a Uber, um LoadBalancer seria interessante, pois imagine que durante certos horários do dia, milhares de pessoas fazem requisições aos seus servidores, enquanto que em outros horários e especialmente de madrugada a demanda é muito menor. Se não houvesse um LoadBalancer, seria necessário ter recursos suficientes para suprir demandas de horários de pico e manter máquinas ociosas durante a noite quando poucas pessoas pedem um Uber.

Concluindo

1) Defina Public Cloud. Cite a principal vantagem e desvantagem.

Public Cloud é um serviço baseado no modelo de computação em nuvem no qual empresas como Amazon (AWS) e Microsoft (Azure) constroem e oferecem recursos computacionais. Esses recursos são máquinas virtuais, aplicações ou servidores de armazenamento, disponíveis para o público geral da internet. Este serviço pode ser oferecido tanto gratuitamente, quanto por um modelo “use e pague”.

Vantagem: possibilitar escalabilidade horizontal para grandes empresas.

Desvantagem: alto custo para implementação.

2) Defina Infrastructure as a Service (Iaas).

Infraestrutura como um Serviço é uma infraestrutura de servidores, rede, armazenamento e outros recursos de computação essenciais que pode ser consumida como um serviço. O objetivo do IaaS é possibilitar a terceirização dessa parte estrutural, assim, o usuário desse serviço pode acessar esses recursos sem ter de se preocupar com manutenção ou com a infraestrutura.

3) Defina Escalabilidade.

No contexto de computação em nuvem, escalabilidade é a capacidade de um sistema de suprir demanda crescente de forma uniforme e estável, com seus recursos atuais de hardware e software. A escalabilidade de um produto pode ser facilmente confundido com sua elasticidade. Um sistema é elástico quando consegue crescer sua oferta de recursos (adicionar mais VMs, por exemplo) conforme um pico de demanda e diminuí-la conforme essa demanda também diminui.

Conclusão: Imagine como é o processo de alocação de uma instância. O que é realmente uma instância? Como você montaria um ambiente similar a AWS em um Datacenter próprio?

O processo de alocação de uma instância envolveria as seguintes etapas:

1. Receber as configurações de instância do usuário.
2. Verificar todas as máquinas daquela categoria e filtrar apenas as que estão disponíveis.
3. Fazer o deploy da AMI para a máquina e iniciar seu monitoramento.
4. Em caso de falha de hardware ou algo do tipo, um backup da AMI é bootado em outra máquina.

Uma instância, então, é uma pequena cópia de um computador, porém essa cópia é virtual e pode-se rodar diversos “computadores virtuais” em apenas uma máquina física.

Para montar um ambiente similar a AWS, podemos utilizar um computador mestre que será responsável por monitorar e fazer os deploys às máquinas. Este computador mestre ficará conectado com os outros e será responsável por gerenciar as máquinas.